

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2003, 20:12:10 ; Search time 3328 Seconds  
(without alignments)

10669.723 Million cell updates/sec

Title: US-09-759-990A-1

Perfect score: 1461  
Sequence: 1 atgcttgcaatcacctac.....ctgatgcttaccgtattataa 1461

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estm:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_hc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_hc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query % Match	Length	ID	Description
1	507.8	34.8	1845	11	AY103583 Zea mays
2	442.2	30.3	1550	12	BM321022 rockefell
3	381	26.1	904	14	CD438088 EL01N0509
4	380.6	26.1	901	14	CD440047 EL01N0532

5	358.2	24.5	800	13	BQ752658
6	343	23.5	1076	12	BG837961
7	339.4	23.2	776	12	BM321423
8	338.8	23.2	843	10	BF256594
9	338.2	23.1	951	12	BM321451
10	335.6	23.0	2111	11	AK075629
11	333.6	22.8	800	12	EL929547
12	333	22.8	774	14	CB682756
13	332.8	22.8	748	13	BQ744296
14	330.2	22.6	977	9	AL163774
15	329.4	22.5	779	14	CB618580
16	329.2	22.5	1163	12	BI244550
17	327.4	22.4	789	14	CB618578
18	325.8	22.3	756	14	CB682413
19	324.8	22.2	724	14	CB655486
20	324	22.2	830	12	BM985729
21	321.4	22.0	709	13	BQ841434
22	319	21.8	711	13	BQ743379
23	318.8	21.8	799	14	CB682388
24	318	21.8	880	14	CB666650
25	318	21.8	936	12	BM816348
26	316.8	21.7	738	14	CD448119
27	316.2	21.6	908	10	BF263419
28	315.2	21.6	716	14	CA452795
29	315	21.6	878	14	CB619506
30	313.6	21.5	867	10	BF253807
31	312	21.4	747	13	BQ993501
32	311	21.3	809	10	BE559309
33	310.2	21.2	698	14	CA932738
34	308	21.1	649	13	BQ295193
35	307.4	21.0	804	12	BM411809
36	306	20.9	693	14	CA928061
37	305.2	20.9	853	14	CB657233
38	305	20.9	682	14	CD440646
39	304.8	20.9	882	13	BQ688034
40	304.2	20.8	682	13	BU672217
41	304	20.8	897	13	BU187457
42	303.8	20.8	977	11	CNS08PT5
43	303.8	20.8	1037	11	CNS08QSM
44	303.6	20.8	727	14	CD437939
45	303.4	20.8	807	14	CA753330

#### ALIGNMENTS

RESULT 1	AY103583	1845 bp	mRNA	linear	HTC 16-OCT-2002
LOCUS	Zea mays PC0113569	mRNA sequence.			
DEFINITION	Zea mays				
ACCESSION	AY103583				
VERSION	AY103583.1	GI:21206661			
KEYWORDS	HTC.				
SOURCE	Zea mays				
ORGANISM	Zea mays				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.				
AUTHORS	Hairney, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.				
TITLE	Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes				
JOURNAL	Unpublished (2002)				
REFERENCE	2 (bases 1 to 1845)				
AUTHORS	Coe, E.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA				
COMMENT	If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at M3L, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the				

maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB:  
www.zmdb.iastate.edu.

## FEATURES

source  
1. 1845  
/organism="Zea mays"  
/mol\_type="mRNA"  
/db\_xref="MaizeDB:636617"  
/db\_xref="taxon:4577"  
/clone\_lib="Maize Mapping Project/DuPont Cornsensus Library"

/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

BASE COUNT 384 a 538 c 514 g 409 t  
ORIGIN

Query Match 34.8%; Score 507.8; DB 11; Length 1845;  
Best Local Similarity 62.0%; Pred. No. 3.6e-134;  
Matches 898; Conservative 0; Mismatches 502; Indels 49; Gaps 4;

```

Qy 34 GAGTACAGAAATGCGACATCAACCTCCATGTTCTCGCGCGTAAGAACTTACCTTGGT 93
Db |||||
Qy 106 GAGTACAAAGTCAAGGATCTCTCGAGCGGACTTCGGCGGCTCGAGATGAGCTGGCC 165
Db |||||
Qy 94 GAGAAAGGAAATGCCAGGTCTTATGGTTCTTGTGAGGGTTATTCGCGCTTTAAAGCCATTG 153
Db |||||
Qy 166 GAGGTCGAAATGCCCGGCTCATGGCGTTCGCGCGCGAGTTCGGCGGTCGACGCCCTTC 225
Db |||||
Qy 154 AAGGTGTCAGATCTCTGGTCCCTCCACATGACATCCAGACAGCGTCTCATCGAG 213
Db |||||
Qy 226 GCGCGCGTAGGATCTCGGGTCTCTCCACATGACCATCCAGACCGCGCTCTCATCGAG 285
Db |||||
Qy 214 ACATTCACAGCTCTGTGTCGATGTCAGATGGGCTTCGTCGAACATCTCTTACACAA 273
Db |||||
Qy 286 ACCCTCACCGGCTCGCGCGGAGTTCGCTGGTCTCTCGAAGATCTTTCGAGGAG 345
Db |||||
Qy 274 GATACAGCGGTGCTGTATCGTTTGGCGCCCAACAGGCAACACAGAGAAGCCGCGGT 333
Db |||||
Qy 346 GACCAGCGCGCGCCATCGCGCGGACTCGGC----- 380
Db |||||
Qy 334 ATCCAGATCTTCGCTGGAGGCGGAAACATCCAGATACCTCCAGATACCTGGAGAAACATACGC 393
Db |||||
Qy 381 --CGCCGTGTTTCGCTGGAGGGGAGACCCCTCGAGGAGTACTGTGTGTCACCGAGCGC 438
Db |||||
Qy 394 GCTCTCACATGGCCAGATGTCGAAGGCCACAGCAGGTTCGTCATGATGTTGTTGATGCT 453
Db |||||
Qy 439 TGCCTTCGACTGGCGGCGGCGCGCCCGGACTCATGTCGACGCGCGGCGGAGCGCC 498
Db |||||
Qy 454 ACATCTCTATCTCCAGGCTTCGAATTCGA-----AACAGCGGTGCTGT 500
Db |||||
Qy 499 AGCTGTCTATCCACGAGGCTGTCAGGCGGAGGAGGATTACGAGGAAGACGCGCAAGAT 558
Db |||||
Qy 501 TCCAGAGCCACAGAACTGACAACTCGAATAACGCTGCTTCGTCGCTTCTGCTACACTCAGCA 560
Db |||||
Qy 559 CCCCAGCCGCGAGTCCACGACAACTGAGTTCAAGATCGTCTCACCATCATCCGGA 618
Db |||||
Qy 561 GGTCTTCAACCAAGCAAGAACCACTGGCACACAGTTCGTCGCGCATGAAGCGTGTTC 620
Db |||||
Qy 619 CGGGCTCAAGCTGACCCCAAGAGGTACCGCAAGTACGAGAGGCTTGTCCGGCTTC 678
Db |||||
Qy 621 CGAAGAGCAACACAGGTTCACCGCTTACCAAGCTCGAGAGGAGGCAACCTCT 680
Db |||||
Qy 679 TGAGAGAGCCACACGCGGTCTCAAGAGGCTCTACCAAGTACGAGAGACCGCGCGCTCT 738
Db |||||
Qy 681 CTTCCAGCATCACTCAAGAGCGTGTACAAAGTCAAGTCCAGTTCATCAATCTAGG 740
Db |||||
Qy 739 CTTCCCTCCCAATTAACGTCAACGATTCCGTCAACAGAGCAAGTTTGACAACTGTATGG 798
Db |||||
Qy 741 CTGCGGCACTCCCTTATCGATGGTATCAACCGTCTTCGATGTCATGATCGCGGCA 800

```

```

Db 799 TTGCGCGCACTCCCTCCCTGATGGTCTGATGAGGCGCACTGACGTATGATGCGCGGAAA 858
Qy |||||
Db 801 GACAGCTCTCGTCACTGGTTACGGCATGTGCGCAAGGCTGCGCTCAATCCCTCCGTGG 860
Qy |||||
Db 859 GTTTCGCGTGTCTGCGGATACGGTGTGTCGCGAGGGTGTGCTGCTCGCTCAAGCA 918
Qy |||||
Db 861 COAAGGCGCTCGGTTATCATCAGAACTCGACCCAAATCTCGCTCTCCAGGCTGCCAT 920
Qy |||||
Db 919 GCTGCTGCGCGTCTCATTTGTGACCGAGATCGACCCCATCTGTGCGCTCCAGGCTCTGAT 978
Qy |||||
Db 921 GGAAGGCTACAGGCTCGCGCATCGAGGAGTCTCAAGGATTCGATATCTTCGTTAC 980
Qy |||||
Db 979 GAGGGTCTTCAGGCTCTTCCTTGGAGGAGTGTCTCTGAAGCTGACATCTTCGTGAC 1038
Qy |||||
Db 981 ATGCACAGGAAATCGGATATCATCTCTGTTGATGATGCGGCCAGATGAAGGATAAGGC 1040
Qy |||||
Db 1039 CACCACCTGGCAACAAGGATATCATGTTGACCATATGAGGAAGATGAAGAACAATGC 1098
Qy |||||
Db 1041 TATTGTCGTAACATCEGCACTTCGATACGAAATTTGATACAGATGCGCTCATGAATA 1100
Qy |||||
Db 1099 CATTTGTCGAAATTTGGCACTTTGACAAATGAATTTGATGCTCGGCTTGAGACCTA 1158
Qy |||||
Db 1101 CCCAGGATCAAGCACATCCCAATCAAGCCAGATACGATCTGSGAATTCGCCAGATGG 1160
Qy |||||
Db 1159 CCTGGCGTCAGCGCATCCATCAAGCCCAAGCTGACCGCTGGGTGTTCCCGGAGAC 1218
Qy |||||
Db 1161 CCAAGCT---ATCCTCTTTTGTGAGGCGCGCTTCTTAACCTTGGCTCGCTCAGG 1217
Qy |||||
Db 1219 CAACACTGGCATCATTTCTTGTGAGGCTGCGCTGATGAGACCTTGGGTGTCTACTGG 1278
Qy |||||
Db 1218 TCACCATCTTTGTTATGATGATTTCAACACAGACACTCGCTCAGCTCGACCT 1277
Qy |||||
Db 1279 CCATCTAGCTTTGTCAATGCTCTCATTTCACTAACCAAGGTCATTGCCCAACTTGAAT 1338
Qy |||||
Db 1278 CTACGAAAGA-----GAGGAATCTCGAAGAAAGTTTACACACTTCGGAAGCATCT 1331
Qy |||||
Db 1339 GTGAGAGAGAGAGCTCTGGCAAGTATGAGAGAGGTTGATGCTGCCCAAGCACT 1398
Qy |||||
Db 1332 CGATGAAGAGTCTGCTCGCTCCACTCGGATCTTCGATGTCACCTTCAAAAGCTTAC 1391
Qy |||||
Db 1399 TGATGAGAGGTTGCTCTCTCCACTTGGGCAAGCTTGGTGCAGAGTGAACCAAGCTCAC 1458
Qy |||||
Db 1392 ACAGAGCAGGCTCACTACATCAAGCTTCCAGTGGGCTCTTCAAGTCTGATGTTA 1451
Qy |||||
Db 1459 CAAGTCTCAGGCGGACTACATCAGGTCGCGATGAGGGTCTCTCAAGCTGCCACTA 1518
Qy |||||
Db 1452 CCGTATTATA 1460
Qy |||||
Db 1519 CCGGTACTA 1527

RESULT 2
BM321022 1550 bp mRNA linear EST 03-JAN-2002
LOCUS rocheFeller 0.1192 Mastigamoeba balamuthi lambda ZAP II Library
DEFINITION Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC 3.3.1.1), mRNA sequence.
ACCESSION BM321022
VERSION BM321022
KEYWORDS EST.
SOURCE Mastigamoeba balamuthi
ORGANISM Mastigamoeba balamuthi
REFERENCE 1 (bases 1 to 1550)
AUTHORS Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W., Gordon,P., Durfle,L., Gaasterland,T., Lopez,P., Muller,M. and Philippe,H.
TITLE The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
MEDLINE 21819461
PUBMED 11830664

```

## COMMENT

Contact: Muller Miklos  
 Laboratory of Biochemical Parasitology  
 The Rockefeller University  
 1230 York Avenue, New York, NY 10021, USA  
 Email: mmuller@rockefeller.edu  
 Insert Length: 1550 Std Error: 0.00  
 POLYA=No.

Location/Qualifiers  
 1. .1550  
 /organism="Mastigamoeba balamuthi"  
 /mol\_type="mRNA"  
 /strain="ATCC 30984"  
 /db\_xref="taxon:108607"  
 /clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"  
 /notes="syn: Phreatamoeba balamuthi"

BASE COUNT 282 a 528 c 495 g 226 t 19 others  
 ORIGIN

Query Match 30.3%; Score 442.2; DB 12; Length 1550;  
 Best Local Similarity 63.4%; Pred. No. 2.2e-115; Indels 10; Gaps 3;  
 Matches 721; Conservative 0; Mismatches 407;

QY 324 GCACGCCGGTATCCAGTCTTCGCTCGAAGGGCGAAACACTCCAGCAATATCTGGGAGAA 383  
 Db 44 GCAGSCCGGGTCTCGGTCTTCGCTCGAAGGGCGAAGCTCCAGGAGTACTGGGAGTG 103

QY 384 CACATACCGGCTCTACATGGCCAGATGTCRAAGGCCACAGCAGGTGTCGATGATGG 443  
 Db 104 CACCTGGAGAGCCCTGTGCTTCGGCCCTTACCAGGGCCCTCAGATCATGTCGACGACGG 163

QY 444 TGGTGATCTACACTCTCTCATCTCCAAAGGGTTCGAAATCGAAACACAGCGGTGCTGTTC 503  
 Db 164 CGGTGACGCACTCTGATGATCCAAAGGGTTCGGCCGCGAGGACACCCCAAGCTGCT 223

QY 504 AGAGCCACAGAGCTGACAACTCGAATACCGTGGCTTCTTGTCTACATCAAGCAGGT 563  
 Db 224 GGAG---GACGACGAGGGCTCGAGAGGTGCGCTGCTCAACAGCGTCTCAAGCAGGT 280

QY 564 CTTCAACCAAGACAGAACCACTGGGCACACAGTGTCTGCGGCATCAAGGTGTTCCGA 623  
 Db 281 CCAGAGGAGCAGCCCGGCTTCTGGCACAAGATCCTCCCGAGATCCGGGTGTCAGCGA 340

QY 624 AGAGACACACAGGTGTCCACGGCTCTACAGCTCGAAGAGGGCAAACTCTCTT 683  
 Db 341 GGAGACGACGCTGCGGTGATGAGGCTGTACAGCTCACCAGCGCAGCAGCTGCTGT 400

QY 684 CCGACCATCAAGTCAACGACGCTGTACAAAGTCCAGTTCGATTAACATCTACGCTG 743  
 Db 401 CCGGCGCTCAAGCTCAAGCACTCTGTCCAAAGAGCAAGTTTGACACATCTACGCTG 460

QY 744 CGGCCACTCCTTATCGATGGTATCAACGGTCTTCGATGTCATGATCGGCGCAAGAC 803  
 Db 461 CGGCCACTTCGCTATCGAGGATCAAGCGCGCACCGAGTGATGCTCGCGGCAAGGT 520

QY 804 AGCTCTCGTATCGGGTATCGGGATGTGGCAAGGGCTGGGCTCAATCCCTCCGTGGCCA 863  
 Db 521 CGCGCTGTCTCGGGCTACGGGACGTGGGACAGGCTGGCCGAGTCTGCTCGCGGCCA 580

QY 864 AGGCGCTCGCTTATCATACAGACTGACCCATCTGGCTCTCCAGGCTGCCATGGA 923  
 Db 581 GGGCTGCGCGCTCATCTGACGGAGATGACCCCATCTGGCGCTCGAGGCTCGATGGC 640

QY 924 AGGCTACAGGTCGCGCGCATCGAGAGTGTCTCAAGGATGTCGATATCTGTTACATG 983  
 Db 641 GGGCTTCAGGTCAACAGCTCGAGCGGGGCTCGACCGCGCGAGATCTCGTGAAGCGG 700

QY 984 CACAGGAACATGCGATATCTCTGTTGACATGATGGCCAGATCAAGGATTAAGGCTAT 1043  
 Db 701 GACGGCAACCAACAACATCATGATGCGCGAGCACAATGACATCGCCCAACTCGGT 760

QY 1044 TGTGGTAAATCGGGCATTGCGATTAACGAAATGATACAGATGGCTCTGATAAATACC 1103  
 Db 761 CGTCTGCAACATCGGCCACTTTTACAAAGAGATCGAGTGGCGGGCTCAGGAAGTGCC 820

QY 1104 AGGCATCAAGACATCCCAATCAAGCCAGAAATACGATGTGGAAATTCGAGATGGCCA 1163  
 Db 821 CGCGCTCAAGTGGTCAACATCAAGCCGAGGTGACAAAGTACGTCTTTGCGCAGGCCA 880

QY 1164 CGCTATCCTCTCTTGTGTGGGGCGGCTTCTTAACCTTGGCTGGCTACAGTCAACC 1223  
 Db 881 GCCATCATCTCTGCTCGCGAGGGCGGCTGTGTCACCCNTCG-TGCGCCACGGGCCACCC 939

QY 1224 ATCTTTGTTATGTCAATGTCAATTCACAAACCCAGACACTCGCTCAGCTTCGACCTCTACGA 1283  
 Db 940 CAGNTTGTCTATGNNACAGCTTCCCAACACAGAGCGTGGCCAGNTNAAAGTTCT---- 995

QY 1284 AAAGAGAGAAATCTCGAGAGAGGTTTACACACTTCGAAAGCATCTCGATGAAGAGT 1343  
 Db 996 --GGCGGAGAAGTGCAGATCGCGGTCAATCACCATCCCAAGATCTCTCGACGAGGAGT 1053

QY 1344 CGCTCGGCTCCACTCGGATCTCTCGATGTCCACCTTACAAAGCTTACAGAGAGCAGGC 1403  
 Db 1054 CGCCGGCTCAGCTCGAGAGCTCAAGTGTCTNTACCCGCTGACCCCGGCCAGGC 1113

QY 1404 TGACTACATCAAGTTCAGTTGAGGTCTCTTAAAGTCTGATGTGTTACCGTTATTA 1461  
 Db 1114 CAAGTACATCGAGTCCCGTGCAGCGCGGTACAAGCGGCCCACTACCGCTANTAA 1171

RESULT 3  
 CD438088 904 bp mRNA linear EST 03-JUN-2003  
 LOCUS EL01N0509B11.b Endosperm\_5 Zea mays cDNA, mRNA sequence.  
 DEFINITION CD438088  
 ACCESSION CD438088  
 VERSION CD438088.1 GI:31353731  
 KEYWORDS EST.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 REFERENCE 1 (bases 1 to 904)  
 AUTHORS Lai, J., Dey, N., Kim, C.S., Beecraft, P., Larkins, B., Linton, E. and  
 Messing, N.  
 TITLE Sequencing of the maize endosperm ESTs  
 JOURNAL Unpublished  
 COMMENT Contact: Lai, Jinheng  
 Dr. Joachim Messing's lab  
 Waksman Institute, Rutgers University  
 190 Frelinghuysen Rd., Piscataway, NJ 08854, USA  
 Tel: 732-445-3801  
 Fax: 732-445-5735  
 Email: jlai@waksman.rutgers.edu  
 Seq primer: T3  
 Location/Qualifiers  
 1. .904  
 /organism="Zea mays"  
 /mol\_type="mRNA"  
 /cultivar="W22"  
 /db\_xref="taxon:4577"  
 /tissue\_type="Endosperm of 7-23DAP"  
 /clone\_lib="Endosperm\_5"  
 /notes="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 204 a 254 c 241 g 205 t  
 ORIGIN

Query Match 26.1%; Score 381; DB 14; Length 904;  
 Best Local Similarity 66.9%; Pred. No. 6e-98;  
 Matches 575; Conservative 0; Mismatches 275; Indels 9; Gaps 2;

QY 612 CGGTGTTTCCGAGAGACACACAGGTGTCCACCGCTCTACAGCTCGAGAGAGGG 671  
 Db 43 CGGCTTTCTGAGAGACACACCGGGTGTCAAGAGGCTCTACAGAGTGCAGGACCGG 102  
 QY 672 CAAACTCTCTTCCAGCCATCAAGTCAACGACGCTGTTCACAAAGTCCAAAGTTCGATAA 731

```

Db 103 GCGCCCTCTCTTCCTGCGCATTAAGCTCAGCATTCGTCACAGAGCAAAATTTGACAA 162
Qy 732 CATCTAAGGCTGCGCCACTCCCTTATCATGATGATGATCAACCGTCTCCGATGTCATGAT 791
Db 163 CTTGTATGTTGCGCGCACTCACTCCCTGATGGTCTGATGAGGGCCACAGACGTTATGAT 222
Qy 792 CGGCGGCAAGACAGCTCTCTGTCATGGTTAGGGGATGCGGAGGCTCGGCTCAATC 851
Db 223 CGCGGTAAAGTTGCGGTGCTGCGGATACGGTGTGTCGCAAGGTTGTCGCGCTGC 282
Qy 852 CTTCCGTGGCCAGAGGCTCGGTTATCATCAGAACTCGACCAATCTGGCTCTCCA 911
Db 283 ACTCAGCAGGCTGGTGGCGGTGTCATGTCACCGATGACACCCACTCTGTCCTCCA 342
Qy 912 GCTGTCATGAAGGCTACAGGTCGCGGATCGAGAGTGTCAAGAGTGTGATAT 971
Db 343 GGCCTGATGAGGGTCTTCAGGTCTTCCTTGGAGGACGTTGTCTGAAGCTGACAT 402
Qy 972 CTTCTTACATGACAGGAACTGGGATATCATCTCTGTTGATGATGCGCCAGATGA 1031
Db 403 CTTCTGACACCACTGGCAAGGATATCATGTTGACCATGAGGAAGATGA 462
Qy 1032 GGATTAAGCTTATGTTGGTAAATCGGCCACTTCGATACGAAATTTGATCAGATGCGCT 1091
Db 463 GAACATGCTATGTTGCAATTTGGCCACTTTTGACATGAATTTGATATGCTCGGCT 522
Qy 1092 CATGAATATCCAGGATACAGACATCCCAATCAGCCAGAAATACGACATGTGGAAAT 1151
Db 523 TGAGACCTTACCTGGGCTCAAGGATCACCATCAAGCCCAAGCTGACCGTGGTGT 582
Qy 1152 CCGAGATGGCAGCT---ATCCTCTCTGCTGAGGGCGCTCTTACCTTGGCTG 1208
Db 583 CCCCAGACCAACTGGCATCATGTCCTTCTGAGGGTGGCTGATGAACCTTGGGTG 642
Qy 1209 CGCTACAGCTCACCCATCTTTGTTATGTCATTCATCAZAAACAGACACTCGTCA 1268
Db 643 TGCTACTGGCATCTAGCTTTGTGATGTCCTGCTCATTCATCAACAGGTCATGGCCA 702
Qy 1269 GCTGCACTCTACGAAGA-----GAGGAATCTCGAAGAAGTTTACACATTC 1322
Db 703 ACTTGAATGTGGAAGGAGAGAGCTCTGGCAAGTATGAGAAGAGGTGATGTGCTCCC 762
Qy 1323 GAAGCATCTGATGAAGAGTGTGCTGCGCTCCACTCGGATCTCTGATGTCACCTTAC 1382
Db 763 CAGACCTTGAAGAGGTTGCTGCTTCCATTTGGGCAAGCTTTGTCGCAAGCTGAC 822
Qy 1383 AAAGCTTACAGAGGAGGCTGACTACATCAACGTTTCCAGTTGAGGGTTCCTTACAAGTC 1442
Db 823 CAAGCTCACCAAGTCTCAGGCGGACTACATCAGGTCGCGATCGAGGTTCCCTACAAGCC 882
Qy 1443 TGATGCTTACGGTTATTA 1461
Db 883 TGCCCACTACGGTACTAA 901

```

## RESULT 4

```

CD440047 CD440047 901 bp mRNA linear EST 03-JUN-2003
LOCUS EL01N0532B10.b Endosperm_5 Zea mays cDNA, mRNA sequence.
DEFINITION CD440047
ACCESSION CD440047
VERSION CD440047.1 GI:31355690
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays

```

```

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1. (bases 1 to 901)

```

```

REFERENCE Lai,J., Dey,N., Kim,C.S., Becraft,P., Larkins,B., Linton,E. and
Messing,J.
TITLE Sequencing of the maize endosperm ESTs
JOURNAL Unpublished

```

## COMMENT

Contact: Lai, Jincheng  
 Dr. Joachim Messing's lab  
 Waksman Institute, Rutgers University  
 190 Frelinghuysen Rd., Piscataway, NJ 08854, USA  
 Tel: 732-445-3801  
 Fax: 732-445-5735  
 Email: jlai@waksman.rutgers.edu  
 Seq primer: T3.

FEATURES  
source

Location/Qualifiers  
 1..901  
 /organism="Zea mays"  
 /mol\_type="mRNA"  
 /cultivar="W22"  
 /db\_xref="taxon:4577"  
 /tissue\_type="Endosperm of 7-23DAP"  
 /clone\_lib="Endosperm\_5"  
 /notes="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 201 a 254 c 242 g 203 t 1 others  
 ORIGIN

Query Match 26.1%; Score 380.6; DB 14; Length 901;  
 Best Local Similarity 65.6%; Pred. No. 7.9e-98;  
 Matches 589; Conservative 0; Mismatches 300; Indels 9; Gaps 2;

```

Qy 571 CAAGACAGAACCACTGCGCACACAGTTGCTCGCGGCATGAACGGTGTTCGGAAGACAA 630
Db 4 CACACGCTCAAGAAGTACCGCAAGATGAAGGAGGCTTTGCGGGCTCTCTGAGGAGACC 63
Qy 631 ACAACAGGTCTCCACCGCTCTTACAGCTCGAAGAGGGGCAAACTCTCTTCCAGGCC 690
Db 64 ACCACGGGTCTCAAGAGGCTCTACAGATCGAGGAGACCGCGCCCTCTCTTCCCTGCC 123
Qy 691 ATCAAGCTCAAGAGCTGTATTACAAAGTCCAAAGTTCGATAACATCTACGGTCGCGCCAC 750
Db 124 ATTAAAGCTCAAGGATTCGCTCACCAGAGCAAGTTTGAACAACCTGTATGTGTGCGGCCAC 183
Qy 751 TCCCTTATCATGATGATCAACCGTCTTCGATGTCATGATCGGGCGCAAGACAGCTCTC 810
Db 184 TCATCTCCCTGATGCTGTATGATGAGGCGCACAGAGCTATGATCGCGGTAGAGTTGCGGTG 243
Qy 811 GTCATGGTTTACGCGGATCTGGCAAGGGTGTGGCTCAATCCCTCCGTGGCAAGGCGCT 870
Db 244 GTCTGCGGATACGGTGTATGTCGCGAAGGTTGTGCGCTGCACTCAAGCAGGCTGTGTGCC 303
Qy 871 CGGTTATCATACAGAACTGACCCCAATCTGCGCTCTCCAGGCTGCCATGGAAGGCTAC 930
Db 304 CGTCTCATTTGACCGAGATCGACCCCATCTGTGCCCTCCAGGCTCTCTGATGAGGGTCTT 363
Qy 931 CAGTCTCGCGCATCGAGGAGTGTCAAGGATGTGATATCTTCGTTACATGCACAGGA 990
Db 364 CAGTCTCTTCCCTTGGAGGACGTTGTCTCTGAAGCTGACATCTTCGTGACCACTGGC 423
Qy 991 AACTGCGATATCATCTCTGTTGACATGATGCGCCAGATGAAGGATAAGGCTTATTGCTGCT 1050
Db 424 AACCAAGGATATCATCATGTTGACCATGAGGAAGATGAAGAACATGCTATGTTCTGC 483
Qy 1051 AACATCGGCCACTTTCGATTAACGAATTTGATACAGATGCGCTCATGAAATACCCAGGATC 1110
Db 484 AACATTCGCCACTTTGACAAATGAAATTTGATATGCTCGGCTTTGAGACCTACCTTGGCGTC 543
Qy 1111 AAGCACATCCCAATCAAGCCAGATACAGATGTGGGAATTTCCAGATGGCCACGCT--- 1167
Db 544 AAGGCGATACCATTAAGCCCCAGACTGACCGGTGGGTGTTCCCGAGACCAACACTGGC 603
Qy 1168 ATCTCCCTTCTGCTGAGGGCGCTTCTTTAACTTTGGCTGGCTCTACAGGTCACCCATCT 1227
Db 604 ATCATTTGCTCTGCTGAGGGTGGCTGATGAACCTTTGGGTGTGCTACTGCGCATCTAGC 663
Qy 1228 TTGTTATGTCATGTCATTCACAAACAGACACTCGCTCAGCTCGACCTCTAGAAAAG 1287
Db 664 TTTGTCTGCTCTGCTCATTTCACTAACCAAGGCTCATTTGCCCAACTTGAACCTGTGAAGGAG 723

```



QY 1288 A-----GAGGAATCTCGAAGAGGTTTACACACTCCGAGCATCTCGATGAGAA 1341  
 |||||  
 Db 724 AAGAGCTCTGCAAGTATGAGAAGAGGTGTATGTGCTCCCAAGCACCITGTATGAGAAG 783  
 |||||  
 QY 1342 GTGCGTCCGCTCCACCTCCGATCTCTCGATGTCCACCTTACAAAGCTTACACAGACAG 1401  
 |||||  
 Db 784 GTTGCTCTCTCCATTGGCGAGCTTGTGTCGCGAGTGTGTCGCGAGTGTGTCGCGAGTGTGTCGAG 843  
 |||||  
 QY 1402 GCTGACTACATCAACGTTCCAGTTCAGGGTCTTACAGTCTGATGCTTACCGTTATT 1459  
 |||||  
 Db 844 GCGGACTACATCAGCTGCGGATCGAGGTCCTTACAAAGCTGCCCACTACCGGTACT 901  
 |||||

## RESULT 5

EQ752658 800 bp mRNA linear EST 23-JUL-2002  
 LOCUS  
 DEFINITION  
 aestivum cDNA clone WHE4117\_F06\_K11, mRNA sequence.

ACCESSION EQ752658  
 VERSION EQ752658.1 GI:21930440  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae  
 ; Triticeae; Triticum.

REFERENCE 1 (bases 1 to 800)  
 Anderson, O.D., Akhunov, E., Chao, S., Crossman, C., Deal, K., Dvorak, J.,  
 Lazo, G.R., Pham, J., Rausch, C.J., Wilson, C. and Woo, J.  
 TITLE The structure and function of the expressed portion of the wheat  
 genomes - Salt-stressed root cDNA library  
 JOURNAL Unpublished  
 COMMENT Contact: Olin Anderson  
 US Department of Agriculture, Agriculture Research Service, Pacific  
 West Area, Western Regional Research Center  
 800 Buchanan Street, Albany, CA 94710, USA  
 Tel: 5105959773  
 Fax: 5105959818  
 Email: oanderson@pw.usda.gov  
 Sequences have been trimmed to remove vector sequence and low  
 quality sequence with phred score less than 20  
 Seq primer: SK primer.

## FEATURES

source  
 1..800  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="Chinese Spring"  
 /db\_xref="taxon:4565"  
 /clones="WHE4117\_F06\_K11"  
 /tissue\_type="Roots"  
 /dev\_stage="Full tillering"  
 /lab\_hosts="E. coli SOLR"  
 /clone\_lib="Wheat salt-stressed root cDNA library"  
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid  
 pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Hydroponic  
 plants grown to full tillering stage were treated with 150  
 mM NaCl for either 12 hours or 7 days. Root tissues of the  
 plants subjected to both types of treatment were collected  
 separately at University of California, Davis (E. Akhunov  
 and K. Deal in J. Dvorak's lab). Total RNA was prepared  
 separately from the two samples (12h and 7day treatments),  
 and equal amount of RNA was then pooled. PolyA RNA was  
 purified from the pooled RNA, a cDNA library was made, and  
 the cDNA clones were in vivo excised to give pBluescript  
 SK(-) phagemids in J. Dvorak's lab (E. Akhunov, J. Dvorak)  
 at the University of California, Davis. Colony plating,  
 plasmid DNA preparations and DNA sequencing were performed  
 in the OD Anderson lab (all other authors)."  
 BASE COUNT 177 a 233 c 210 g 180 t  
 ORIGIN

Query Match 24.6%; Score 359.2; DB 13; Length 800;  
 Best Local Similarity 67.6%; Pred. No. 2e-91;

Matches 536; Conservative 0; Mismatches 248; Indels 9; Gaps 2;  
 QY 661 GAGAAGGAGGGCAAACTCTCTTCCCAGCCATCAACGTCACAGCGTGTTTACAAAGTCC 720  
 |||||  
 Db 7 GACGAGGCGGGACCCCTCTCTTCCCAGCCATCAACGTCACAGCTCGTCCACAGAGC 66  
 |||||  
 QY 721 AAGTTCGATTAACATCTACGGCTGCGGCACCTCCCTTATCGATGGTATCAACCGTGTCTCC 780  
 |||||  
 Db 67 AAGTTTACAACTTTACGGTGTGCGGTCACTCGCTCCCTGATGSGTCTTATGAGGGCCACT 126  
 |||||  
 QY 781 GATCTCATGATCGGGGCAAGACAGCTCTCGTCAATGGGTTACGGGATGTCGGCAAGGGC 840  
 |||||  
 Db 127 CATGTATGATCGCGGCAAGGTTGCCGTGTTATGGGTATGGTATGGCAAGGGC 186  
 |||||  
 QY 841 TGCCTCAATCCCTCGTGGGCAAGGCGCTCGGTTATCATCACAGAACTCGACCAATC 900  
 |||||  
 Db 187 TGTGCCGCCGCACTTAAAGCAGGCTGTGCGCGTGTGATCGTACAGAGATTGACCCCAT 246  
 |||||  
 QY 901 TGCCTCTCCAGGCTGCCATGGAAGGCTACAGGTGCGGCGCATCGAGGAAGTCTCAAG 960  
 |||||  
 Db 247 TGTGCTCTTCAGGCCCTGATGGAGGGATCCAGATCTCACCTTGGAGGATGTTGTCTCT 306  
 |||||  
 QY 961 CATGTCCATATCTTCTGTACATGCACAGGAACTGGGATATCATCTCTGTGACATGATG 1020  
 |||||  
 Db 307 GAGCTGATATCTTGTGACACACACCGGAAACAGAGACATCATCATGTTGACCAATG 366  
 |||||  
 QY 1021 GCCAGATGAAGATAAGGCTATTGTGCGTAACATCGGCCACTTCGATAACGAATTGAT 1080  
 |||||  
 Db 367 AGGAAGATGAAGAACAACGCCCATTTGTCTGCAACATTTGGTCACTTTTGACACAGAGATTGAC 426  
 |||||  
 QY 1081 ACAGATCGCTCATGAAATACCCAGGATCAAGCACATCCCAATCAAGCCAGAGATACGAC 1140  
 |||||  
 Db 427 ATGAACGGCTTTGAGACCTTACCTCTGTTCAAGCGCATCACATCAAGCCCGACACTGAC 486  
 |||||  
 QY 1141 ATGTGGGAATTCACAGATGGCCA---CGCTATCTCTCTTCTTGTGAGGGCGCGCTTCTT 1197  
 |||||  
 Db 487 CGCTGGGTCTTCCCGAGACCAAGACCGGCATCATTTGTTCTTGTGAGGGTCTGCTGATG 546  
 |||||  
 QY 1198 AACCTTGGCTGGCTACAGGTACCGCATCTTTCGTTATGTCAATGTCTATTCACAAACCAG 1257  
 |||||  
 Db 547 AACCTTGGATGTGCCACTGGCCACCCAGCTTTGTCTCATGTCTCTTCACTTCACTTAACAG 606  
 |||||  
 QY 1258 ACATCTCGCTCAGCTCGACCTCTACGAAAG-----AGAGGAAATCTCGAGAGAGGTT 1311  
 |||||  
 Db 607 GTTATTTCTCAGCTTGAGTTGTGGAACGAGAGGCGCACCGCAAGTACGAGAGAGGTG 666  
 |||||  
 QY 1312 TACACACTTCCGAAGCATCTCGATGAAGAAAGTCGCTCGCTCCACTCGGATCTCTCGAT 1371  
 |||||  
 Db 667 TACGTTCTTCCCAAGCACCTCGAGAGAGAGGTCTCGGCGCTCCACTTGGGCAAGCTCGGC 726  
 |||||  
 QY 1372 GTCCACCTTACAAAGCTTACACAGAGGAGGCTGACTATCATCAAGTTTCCAGTTGAGGTT 1431  
 |||||  
 Db 727 GCCAGGCTGACCAAGCTCACCAAGTCCAGTCTGACTATTAGCATCCCAAGTTGAGGTT 786  
 |||||  
 QY 1432 CTTTACAAAGTCTG 1444  
 |||||  
 Db 787 CTTTACAAAGCTG 799  
 |||||

RESULT 6  
 EQ837961  
 LOCUS  
 DEFINITION  
 Zm10\_03e04 A Zm10\_AAPC\_ECORC\_Fusarium graminearum corn\_silk Zea  
 mayis cDNA Clone Zm10\_03e04, mRNA sequence.  
 ACCESSION EQ837961  
 VERSION EQ837961.1 GI:14204284  
 KEYWORDS EST.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 REFERENCE 1 (bases 1 to 1076)

**AUTHORS** Harris,L.J., Balcerzak,M., Allard,S., Saparno,A., Couroux,P., De Moors,A., Hattori,J.I., Ouellet,T., Robert,L.S., Singh,J.A, Sprott ,D. and Finker,N.A.  
**TITLE** Expressed Sequence Tags from Maize Silk Six Hours After Silk Channel Inoculation with Fusarium graminearum  
**JOURNAL** Unpublished  
**COMMENT** Contact: Harris, Linda J.  
 Eastern Cereal and Oilseed Research Centre  
 Agriculture and Agri-food Canada  
 Bldg. 21, Central Experimental Farm, Ottawa, Ontario, KIA 0C6, CANADA  
 Tel: (613) 759-1314  
 Fax: (613) 759-6566  
 Email: harrisjl@em.agr.ca.

**FEATURES**  
 source  
 1..1076  
 /organism="Zea mays"  
 /mol\_type="mRNA"  
 /culivar="CO388"  
 /db\_xref="taxon:4577"  
 /clone="Zm10.03e04"  
 /tissue\_type="Silk"  
 /dev\_stage="4-5 days post-silk emergence"  
 /clone\_lib="Zm10\_AAFRC\_EOORC\_Fusarium\_graminearum\_corn\_silk"

"/note="Vector: Bluescript SK+/XhoI-EcoRI; Site 1: EcoRI; Site2: XhoI; Field-grown corn was silk channel-inoculated in the morning (~10 am) with 1 ml of a macroconidial suspension (500,000 spores/ml) of Fusarium graminearum and silk channels were collected and immediately frozen in liquid nitrogen 6 hours later. RNA was extracted from silk tissue between 1 cm below and above the inoculation point in the silk channel. RNA from five silk channels was pooled."  
**BASE COUNT** 228 a 281 c 288 g 278 t 1 others  
**ORIGIN**  
 Query Match 23.5%; Score 343; DB 12; Length 1076;  
 Best Local Similarity 67.2%; Pred. No. 5.6e-86;  
 Matches 533; Conservative 0; Mismatches 250; Indels 10; Gaps 3;

QY 678 CTTCTTCCCGCCATCAAGCTCAAGCGCTGTATACAAAGTCCAAAGTTCGATAACACTA 737  
 Db 1 CTTCTTCCCGCCATTAAGTCAAGCTCAAGCTTCGTCACCAAGACCAAGTTTGACACTGTA 60  
 QY 738 CGCTGCGGCACATCCCTTATCGATGATGATCAACCGTGTTCGATGTCGATGTCGCGG 797  
 Db 61 TGGTTGCGGCACACTCACTCCCTGATGATGATGATGATGATGATGATGATGATGATGAT 120  
 QY 798 CAAGACAGCTCTGTCATGGTTACGGATGTCGGATGTCGGAGGCTGGCTCAATCCCTCCG 857  
 Db 121 TAAGTTGCGTGTCTCGGATACGGTATGTTGGCAAGGGTGTGCGGTGCGTCACTCAA 180  
 QY 858 TG-GCCAGGCGCTCGGTTATCATCAGAACTCGACCAATCTCGCTCTCCAGGCTG 916  
 Db 181 RCGAGCTGTGTCGCGGTGTCATTGTGACCGAGATCGACCCCACTCGCGCTCCAGGCTC 240  
 QY 917 CCATCGAAGCTACAGGTCGCGCCGATCGAGGAGTGTCAAGATGTCGATATCTTCG 976  
 Db 241 TGATGGAGGTTCTTCAGGTCCTTCCCTTGGAGGAGCTGTCTCGGAAGCTGACATCTTCG 300  
 QY 977 TTATATGCACAGGAACCTCGATATCATCTCTGTGATGATGATGCCACATGAAGGATA 1036  
 Db 301 TGACACCACTGGCACACAGGATATCATCATGTTTGACCATGAGGAGATGAAGACA 360  
 QY 1037 AGGCTATTGTGTTAATCATCGGCCACTTCGATTAACGAAATTTGATACAGATGGCTCATGA 1096  
 Db 361 ATGCCATTGTCTCAACANTGGCCACTTTGACATGAATTTGATATGCTCGGCTTCAGA 420  
 QY 1097 AATACCCAGGATCAAGACATCCCAATCAAGCCAGATACAGCATGTGGGAATTCCTCCAG 1156  
 Db 421 CTTACCTTGGCGTCAAGCGCATCACCATCAAGCCCAAGATGACCGCTGGGTGTTCCCGC 480

QY 1157 ATGCCACGCT--ATCCTCCTTTTCTGCTGAGGCGCGCTTCTTAACTTGGCTGCGCTA 1213  
 Db 481 AGACCAACACTGGCATCATTTGTCTTCTGCTGAGGTGCGCTGATGAACCTTGGTGTGCTA 540  
 QY 1214 CAGTTCACCCATCTTTCTGTTATGTCATATGTCATTAACCAACAGACACTGCTCAGCTCG 1273  
 Db 541 CTGGCCATCTCCTAGCTTTGTCAITGCTGCTCATTAACCAAGGTCAATGCCCACTTG 600  
 QY 1274 ACCTTCTACGAAAGA-----GAGGAAATCTCGAAGAAGGTTTACACACTTCCGAAGC 1327  
 Db 601 AACTGTGGAAGGAGAAGAGCTCTGCAAGTATGAGAAGAAGGTGATGTGCTCCCAAGC 660  
 QY 1328 ATCTCGATGAAGAGTCGCTCGCTCCACCTCGATCTCTCGATGTCACCTTACAAAGC 1387  
 Db 661 ACCTTGTATGAGAAGGTGCTGCTCTCCACTTGGCGAAGCTTGTGCGCAAGTGACCAAGC 720  
 QY 1388 TTACACAGAAAGAGGCTGACTACATCAACGTTCCAGTTCAGGTCCTTACAAAGTCTGATG 1447  
 Db 721 TCACCAAGTCTAGGCCGACTACATCAGCGTGCAGATCGAGGTCCTTACAAAGCTTGCCC 780  
 QY 1448 CTTACGCTTATTA 1460  
 Db 781 ACTACCGGTACTA 793

**RESULT 7**  
**LOCUS** BM321423 776 bp mRNA linear EST 03-JAN-2002  
**DEFINITION** rockefeller.0.1256 Mastigamoeba balamuthi lambda ZAP II Library  
 Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC 3.3.1.1), mRNA sequence.  
**ACCESSION** BM321423  
**VERSION** BM321423.1 GI:18055829  
**KEYWORDS** EST.  
**SOURCE** Mastigamoeba balamuthi  
**ORGANISM** Mastigamoeba balamuthi  
 Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.  
**REFERENCE** 1 (bases 1 to 776)  
**AUTHORS** Baptiste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W., Gordon,P., Druffe,L., Gaasterland,T., Lopez,P., Muller,M. and Philippe,H.  
**TITLE** The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba  
**JOURNAL** Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)  
**MEDLINE** 21819461  
**PUBMED** 11830664  
**COMMENT** Contact: Muller Miklos  
 Laboratory of Biochemical Parasitology  
 The Rockefeller University  
 1230 York Avenue, New York, NY 10021, USA  
 Email: mmuller@rockvax.rockefeller.edu  
 Insert Length: 776 Std Error: 0.00  
 POLYA=No.

**FEATURES**  
 Location/Qualifiers  
 1..776  
 /organism="Mastigamoeba balamuthi"  
 /mol\_type="mRNA"  
 /strain="ATCC 30984"  
 /db\_xref="taxon:108607"  
 /clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"  
 /note="syn: Phreatamoeba balamuthi"

**BASE COUNT** 158 a 266 c 232 g 113 t 7 others  
**ORIGIN**  
 Query Match 23.2%; Score 339.4; DB 12; Length 776;  
 Best Local Similarity 67.2%; Pred. No. 4.9e-86;  
 Matches 503; Conservative 0; Mismatches 243; Indels 2; Gaps 2;

QY 553 CTCAGCAGGTCTTCAACCAAGACAAACCACTGGCACACAGTTGTGCGCGCATGAAC 612  
 Db 12 CTCAGCAGGTCTCAGAAAGAGCAGCCCGGTTCTTGGCAAGATCTCTCCCGAGATCCGC 71  
 QY 613 GGTCTTCCGAAGAGACAAACAGGTGTCTCCACCGCTCTTACAGCTCGAAGAGGAGC 672

```

Db      72  GGTGTACGAGGAGAGACGACTGCGNNNNNNNGTGTACAGTGCACCGCAGCGC 131
QY      673  AAACTCTCTTCCAGCCATCAACGTCAACGACGCTGTACAAAGTCCAAAGTTCGATAAC 732
Db      132  AAGCTGTGTTCGCGCGCTCAACGTCAACGACTCTGTCCACAGAGCAAGTTCGACAAC 191
QY      733  ATCTACGCGTCCGCGCAGCTCCCTTATCGATGGTATCAACCGTGTTCGATGTCAATC 792
Db      192  ATCTACGCGTCCGCGCAGCTCGTCTATCGACGCGCATCAAGCGCGCACGAGTGTATCTC 251
QY      793  GCGCGCAAGACAGCTCTCTCATGTGGTTACGCGCATCGGCAAGGCTGGCTCATATCC 852
Db      252  GCGCGCAAGTTCGCGCTGTTCGCGGCTACGCGCAGCTGGCAAGGCTGGCGGAGTCG 311
QY      853  TCCGCTGCCAAGCGCTCGCTTATCATCAAGAACTCGACCCCAATCTGGGCTCTCCAG 912
Db      312  TCGCGCGCCAGGCTCGCGCTCATCTGACGAGATCGACCCCATCTGGCGCTGCAG 371
QY      913  GTCGCGATGAGGCTACAGGCTCGCGCATCGAGGAAGTGTCAAGATGTCAATC 972
Db      372  GCGTCGATGCGCGGCTTCGAGTCAACACGCTCGAGCGGGCTCGACCGCGCGAGATC 431
QY      973  TTCGTTACATGCACAGAACTCGCATATCATCTCTGTTGACATGATGCCCGCATGAAG 1032
Db      432  TTCGTGCGCGGAGCGGCGACACACAACTATCATATGCGCGAGCATGAGCAAGTCGC 491
QY      1033  GATGAAGCTATTTCGTTAACTCGGCCACTTGTGATAACGAAATGATACAGATGGCCTC 1092
Db      492  CACAACCTCGGTCTGCAACATCGGCCACTTTGACAAAGATCGAGTTCGCGGCTC 551
QY      1093  ATGAATACCCAGGATCAAGACATCCCATCAAGCAGATAGCATGTGGAAATC 1152
Db      552  AAGAGTGGCGCGGCTCAAGTGCCTCAACATCAAGCC-GCACTCGAAGATAGTCTTT 610
QY      1153  CCAGATGCCACGCTATCTCTTCTGCTGAGGCGCGCTTCTTAACTTGGCTGGCT 1212
Db      611  GCGAGCGCCACCGCCATCTCTGCTCGGAGGCGCGCTGCTCACTCGGCTGGC- 669
QY      1213  ACAGGTACCCATCTTTCGTTATGTCATGTCATCAAAACAGACACTCGCTCAGCTC 1272
Db      670  ACGGCGCCACCCAGCTTCGTCATGAGCACCAGCTTCACCAACAGACGCTGGCCAGATC 729
QY      1273  GACCTCTACGAAGAGAGAAATCTCG 1300
Db      730  AAGCTCTGCGCGAGAAAGTSCGACATCG 757

```

## RESULT 8

BF256594

LOCUS

DEFINITION BF256594 843 bp mRNA linear EST 22-OCT-2001  
 (Etiolated and unstressed) Hordeum vulgare subsp. vulgare cDNA  
 clone HVSMF0010H04f, mRNA sequence.

ACCESSION

BF256594

VERSION

BF256594.2

KEYWORDS

EST

SOURCE

Hordeum vulgare subsp. vulgare

ORGANISM

Hordeum vulgare subsp. vulgare

REFERENCE

1. (bases 1 to 843)

AUTHORS

Wing, R., Close, T.J., Klein, D., Wise, R., Begum, D., Frisch, D., Yu

, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton

, R.D., Oates, R., and Main, D.

TITLE

Development of a genetically and physically anchored EST resource

JOURNAL

for barley genomics: Morex unstressed seedling root cDNA library

COMMENT

Unpublished

On Nov 16, 2000 this sequence version replaced gi:11185707.

Contact: Wing RA

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu  
 Total hg bases = 615  
 Seq primer: AATTAACTCTCACTAAAGGG  
 High quality sequence stop: 714.

## FEATURES

source

1. .843

Location/Qualifiers

/organism="Hordeum vulgare subsp. vulgare"

/mol\_type="mRNA"

/cultivar="Morex"

/db\_xref="taxon:112509"

/clone="HVSMF0010H04f"

/tissue\_type="Seedling root"

/lab\_host="TJG121"

/clone\_lib="Hordeum vulgare seedling root EST library

/note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;

Seeds were surface sterilized then germinated under axenic

conditions in the dark at room temperature on filter paper

with water, nystatin and cefotaxime in covered

crystallization dishes. Five-day old seedling roots were

then harvested, total RNA was prepared, poly(A) RNA was

purified, one primary unamplified cDNA library was made,

and 1 million pfu were in vivo excised to give plasmids

SKI(-) cDNA phagemids. These steps were performed in the TJ

Close laboratory at the University of California,

Riverside (Choi, Close, Fenton). Phagemids were plated and

picked at the Clemson University Genomics Institute (CUGI)

(Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA

preparations, DNA sequencing and sequence analysis were

performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates

, Rambo, Main). The sequence has been trimmed to remove

vector sequence and contains a minimum of 100 bases of

phred value 20 or above. For more details on library

preparation and sequence analysis see

http://www.genome.clemson.edu/projects/barley. To order

this clone see http://www.genome.clemson.edu/orders Also

see Close TJ, Wing R, Klein, D., Wise R (2001)

Genetically and physically anchored EST resources for

barley genomics. Barley Genetics Newsletter 31:29-30.

(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

BASE COUNT 192 a 251 c 220 g 179 t

ORIGIN

Query Match 23.2%; Score 338.8; DB 10; Length 843;

Best Local Similarity 64.5%; Pred. No. 7.7e-86;

Matches 540; Conservative 0; Mismatches 288; Indels 9; Gaps 2;

QY 518 CTGACACCTCGAATACCGCTGCTTCTGCTACACTCAAGCAGGTCTTCAACCAAGACA 577

Db 7 CCGACACCCCGAGTGAAGATGCTCTCCACATCATCGGAGCGCTCAAGACCGAGC 66

QY 578 AGAACCACTGGCACACAGTGTCTCGCGCATGAACGGTGTTCGGAAGAGACAAACAG 637

Db 67 CCAGCAAGTACCGCAAGATGAGAGAGGCTCTCGGTGTCTCCAGGAGACCAACCG 126

QY 638 GTGTCCACCGCTTACAGCTCGAGAAGAGGCGGCAAACTCTCTTCCAGGACCAACG 697

Db 127 GCGTCAAGAGGCTTACAGATGCGAGGTCGCGGACCCCTCTCTTCCGCCCATCAAG 186

QY 698 TCAACGACGCTTACAAGTCCAGTTGATGAATCATCTAGCGTGGCGGCTCTCCCTTA 757

Db 187 TCACGACCTCCGTCACCAAGAGCAAGTTTGACAACTTTACGGTTGGCGTCACTCTCC 246

QY 758 TCGATGATCAACCGTCTCTCCGATGTCTATGATGGGCAAGACAGCTCTCGTCATCG 817

Db 247 CTGACGCTCTCATAGGCGCACTGATGTATGATCCCGTAAAGTCCCGTGTCTCGG 306

QY 818 GTTACGCGGATCTCGGCAAGGCTGCGCTCAATCCCTCCGTGGGCAAGGCGCTCGGCTTA 877

Db 307 GTTATGCTGATGTGGCAAGGCTGTGCGCGCGCATCTCAAGCAGGCTGGTGGCGGTGCA 366

```

QY 878 TCATCAGAACTCGACCCAAATCTGGCTCTCCAGGTGCGATGGAAGGCTACCAAGTCC 937
Db 367 TCGTGACAGAGATCGACCCCAATCTGGCTCTCCAGGCTGATGAGGGCATCCAGATCC 426
QY 938 GCGGATCGAGAAAGTGTCAAGATGTCGATATCTTCGTTATCATGCAAGAACTGGCG 997
Db 427 TCACCTTGAGGATGTGTCTCGAGGCTGATATCTTGTGACCAACCCACCGAAGG 486
QY 998 ATATCATCTCTGTGTGATGATGCGCCAGATGAAGGATAGGCTATTTGGTGAACATCG 1057
Db 487 ACATCATCATGTGTGACCATGAGGAAGATGAGAAACAGCCATTTGTGCAACATTG 546
QY 1058 GCCACTTCGATTAAGAAATGTATACAGATGCGCTCATGAATAACCGAGGATCAAGACA 1117
Db 547 GTCACTTTCACACAGATCGACATGAAACGGCTTGAGACCTACCCCGGTGTGATCGGCA 606
QY 1118 TCCCAATCAAGCCAGATAGCATGTGGAAATTCAGATGGCCA---CGCTATCTCC 1174
Db 607 TCACCATCAAGCCCAAGATGACCGTGGGTCTTCCCTGAGACCAAGACCGGCATATTG 666
QY 1175 TTCTTGTGAGGCGCGCTTCTTAACCTTGGCTGCGTACAGGTCAACCATCTTTTGGTTA 1234
Db 667 TTCTTGTGAGGCTGTCTGTGATGAACCTTGGATGTGCCACTGGCCACCCAGCTTTGTCA 726
QY 1235 TGTGAATGTATTAACAAACAGACACTGCTGCTGACCTGCTGACCT-----CTACGAAAGA 1288
Db 727 TGTNCTGTCTTTTCACTAACCAAGGTTATTGCCAGCTCGAGTTGTGAAACGAAGGCCA 786
QY 1289 GAGGAAATCTCGAAGAGGTTTACACACTTCCGAAGCATCTCGATGAAGAAGTCG 1345
Db 787 GCGGCAAGTACGAAAGAGTGTGTCTTCTCCAGACCTCGACAGAGGTTG 843

RESULT 9
LOCUS BM321451
DEFINITION rockefeller 0.1211 Mastigamoeba balamuthi lambda ZAP II Library
Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC
3.3.1.1), mRNA sequence.
BM321451
VERSION BM321451.1 GI:18055857
KEYWORDS EST.
SOURCE Mastigamoeba balamuthi
ORGANISM Mastigamoeba balamuthi
REFERENCE 1 (bases 1 to 951)
AUTHORS Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
Philippe,H.
TITLE The analysis of 100 genes supports the grouping of three highly
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
21819461
11830664
CONTACT Muller Miklos
LABORATORY of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
EMAIL: mmuller@rockvax.rockefeller.edu
INSERT LENGTH: 951 Std Error: 0.00
POLYA=No.
FEATURES
1. .951
Location/Qualifiers
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/db_xref="taxon:30984"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
/note="syn: Phreatamoeba balamuthi"
BASE COUNT 186 a 321 c 303 g 139 t 2 others
ORIGIN

```

```

Query Match 23.1%; Score 338.2; DB 12; Length 951;
Best Local Similarity 62.2%; Pred. No. 1.2e-85;
Matches 598; Conservative 0; Mismatches 330; Indels 33; Gaps 3;

QY 33 CGAGTACAGAAATGCGGACATCAACCTCCATGTTCTCGGCGTAAAGAACTTACCCCTGC 92
Db 24 CGACTCAAGGTCAGAGACATCAACCTTCGCCGACTTCGGCGCAAGGAGATCGAGTCGC 83
QY 93 TGAAGAGAAATGCCAGGCTTTATGGTTCTTCGTGAGGCTTATTCCCGCTTCTAAGCAAT 152
Db 84 CGAGTACAGAGATGCCGGTCTGTATGGCCACAG---GACCAAGTACGCCGCCGAGTGCT 140
QY 153 GAAGSGTGTCAAGATCTCTGGTTCCCTCCATGACAGTCCAGACAGCCGCTCTCATCGA 212
Db 141 CCAGGGCGGAGGATCAACCGGCTCGCTCCACATGACGATCCAGACCGCGCTCTCATCGA 200
QY 213 GACATCAACAGCTCTTGGTGTGATGTCAGATGGGCTTCCTGCAACATCTTCTCTACACA 272
Db 201 GACCTCAAGTTCTTCGGGCCAAGATCCGCTGGTGTCTGTGCAACATCTTCTCGACGA 260
QY 273 AGATPACAGCGCTGTGCTATGTTGTGCGGCCAACAGGACACACAGAGAAAGCCAGCGG 332
Db 261 GGACCAAGCGGCGCGGCCCATCG-----CGCAGGCGCG 293
QY 333 TATCCAGAGTCTTCGCTCGAAGGCGGAAACATCTCCAGAACTACTGGGAGAACACATACCG 392
Db 294 CGTCTCGGTTCTCGCTCGAAGGCGGAGAACCTCCAGGAGTACTGGGAGTGACCTTGGAA 353
QY 393 CGCTTCTCATATGCGCAGATGTTCAAGGCCACACAGCAGGTTGTGATGATGATGTTGATGC 452
Db 354 GGCCCTTGTCTCGGCCCTTACCAGGGCCCTCAGATCATCTGTCGACGAGCGGCTGAOCG 413
QY 453 TACATCTCTCATCTCGAAGGCTTCGAATTCGAAACAGCGGCTGTCTTCCAGAGCCAC 512
Db 414 GACTCTGATGATCCCAAGGGGTTTCGCGCGCGAGGACAAACCCCAAGTGTGTGAG---GA 470
QY 513 AGAAGCTGACCAACTCGAATACCGCTCGCTTCTGTCATACATCAAGCAGGTCTTCAACA 572
Db 471 CGACAGGGCTCGAGGAGTTCGCTGCTCCATCAACACGTGTCACAGCAGGTCCAGAGGA 530
QY 573 AGACAAGAAACCACTGGCAACAGTTGTGTCGCGCATGAAAGGTTTTCGAAAGAGCACAC 632
Db 531 CGAGCCCGGCTTCTGGCAACAGATCTCCCGGAGATCCCGGTGTGTCAGCGAGGAGACGAC 590
QY 633 AACAGGTGTCACCGCTCTACAGGCTCGAAGAGGAGGCAAACTCCTCTTCCAGCCAT 692
Db 591 GACTGGCGTATGAGGCTGTACAGCTGCACCGGAGCGCAGAGTGTCTGTTCGCGCGCT 650
QY 693 CAAGTCAACGAGCTGTTCACAAAGTCCAAAGTTCGATTAACATCTACGCTGCGCGCACTC 752
Db 651 CAAGTCAACGACTCTTNTCCAAAGAGCAAGTTTCGACACATCTACGCTGCGCGCACTC 710
QY 753 CTTATCGATGATACACCGGCTTCGATGTCATGATCGCGGCGCAAGACAGTCTCTGT 812
Db 711 GCTCATCGAGGCAATCAAGCGCGCACCGCTGATGCTCGCGCGGCAAGGTCGCTGT 770
QY 813 CATGGGTTACGGCGATGTGCGCAAGGCTCGGCTCAATCCCTCGGTGCGCAAGCGCTCG 872
Db 771 CGCGGCTACCGCGAGCTGGGCGAAGGCTGCGCGGAGTCTGCTGCGCGGCAAGGCTCGCG 830
QY 873 CTTATCATCAGAACTCGAACCAATCTCGCTTCCAGGCTGCCAGTGCATGGAAGGCTACCA 932
Db 831 CGTCATCGTACGAGATCGACCCCATCTCGCGGCTGACAGCGTGCATGCGCGGCTTCCA 890
QY 933 GGTCCGCGCATCGAGGAGTTCGCAAGGATGTCGATATCTTCGTATCATGCAAGGAA 992
Db 891 GGTCAAACGCTCGAGGCGGGGCTCGAGCCGCGCGAGATCTTCTGTGAGCGGCGAGGCAA 950
QY 993 C 993
Db 951 C 951

```

RESULT 10	AK075629	2111 bp	mRNA	linear	HTC 07-DEC-2002
LOCUS	AK075629				
DEFINITION	Mus musculus 18-day embryo whole body cDNA, RIKEN full-length enriched library, clone:1110019K01 product:S-adenosylhomocysteine hydrolase, full insert sequence.				
ACCESSION	AK075629.1	GI:26344432			
VERSION	1				
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 Carninci,P. and Hayashizaki,Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2				
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				
REFERENCE	3				
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,K., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujikawa,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer				
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)				
MEDLINE	20530913				
PUBMED	11076861				
REFERENCE	4				
AUTHORS	Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojohori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Stauble,J., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldairelli,R., Barsh,G., Blake,J., Soffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawai,J.H., Kottuski,S. and Hayashizaki,Y.				
TITLE	Functional annotation of a full-length mouse cDNA collection				
JOURNAL	Nature 409 (6821), 685-690 (2001)				
MEDLINE	21085660				
PUBMED	11217851				
REFERENCE	5				
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.				
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs				
JOURNAL	Nature 420, 563-573 (2002)				
MEDLINE	6 (bases 1 to 2111)				
PUBMED	Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,				
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL</					

```

Db      266  GAGTCCCTCATAGATGCGATCAAAACGGGCCACAGATGTGATGATTCGGGCAAGGTGGCG 325
QY      808  CTGCTCATGGGTTACGGCGATTCGGCAAGGGCTCGCTCAATCCCTCGTGGCCAAAGGC 867
Db      326  GTGTGGCAGGCTATGCTGATGTGGCAAGGCTGTGCCAGGCCCTGAGGGGTTTGGG 385
QY      868  GCTCGGTTATCATCAGAACTCGACCCAAATCTGGCTCTCCAGGCTGCCATGGAAGGC 927
Db      386  GCGCGAGTCATCATCCGAGATCGACCCCATCAATGCACTGCAAGCTGCCATGGAGGC 445
QY      928  TACCAGGTCCGCGCATCGAGAAGTCTGCAAGATGTGATATCTTCGTTACATGCACA 987
Db      446  TATGAGGTACCCTATGACGAGACCTGTAGAGGGGCAACATTTTGTCCACCACA 505
QY      988  GAAACTCCGATATCATCTCTGTGACATGATGGCCAGATGAAGATAAGCTATTGTC 1047
Db      506  GCGTGTGGATATCATCTTGGCGGCACCTTTGAGCAGATGAAGATGAGCCATTGTC 565
QY      1048  GGTAACTCGGCCACTTCGATTAACGAATTCATACAGATGGCTCATGAATACCCAGGC 1107
Db      566  TGTAACTTGGACACTTCGATGTGAGATTTGATGTAAGTGGCTCAATGGAACGGCG-- 623
QY      1108  ATCAAGCACATCCCAATCAAGCCAGATACGACATGTGGGAATTCCCAGATGGCCAGCT 1167
Db      624  -TGGAGAAAGTGAACATCAAGCCCAAGTGGACCGCTACTGGCTAAAGATGGCGCGC 682
QY      1168  ATCTCTCTTCTGCTGAGGCGCGCTTCTTAACTTGGCTGGCTGACAGGTCAACCATCT 1227
Db      683  ATCATCTTGTGCTGAAGCGCGCTCTGCTCAACCTGGGTGTGCCATGGGACACCCAGC 742
QY      1228  TTCGTTATGTCAATGTCAATCAAAACAGACACTGCTCAGCTGAGCTCTACGAAAG 1287
Db      743  TTCGTGATGACACTCTCTTCAAAACCGAGTGTGACAGATGAGCTGTGACCCAC 802
QY      1288  AGAGAAATTCGAGAAGAAGTTTACACACTTCGGAAGCATCTCGATGAAGAAGTGGCT 1347
Db      803  CCAGATAAATACCTGCTGGGGTTCACTTCTCGCTTAAGAAGCTGGATGAGCGGTGGCT 862
QY      1348  CGCTTCACTTCGATCTCGATGTCCACCTTCAAAAGTTTACAGAGCTGACAGAGCGGTGAC 1407
Db      863  GAAGCCCACTGGGCAAGCTGATGTGAAGCTGACCAAGCTGACTGAGAAGCAGGCCAG 922
QY      1408  TACATCAACGTTCCAGTTGAGGCTCTTTACAAGTCTGATGCTTACCGTTATTAA 1461
Db      923  TACCTGGCATGCCCATCAAGGCCCTTTCAAGCTTGATCATACCGGTACTGA 976

RESULT 11
BI929547      800 bp      mRNA      linear      EST 18-OCT-2001
LOCUS      EST549436 tomato flower, 3 - 8 mm buds Lycopersicon esculentum cDNA
DEFINITION      clone cTOB29E10 5' end, mRNA sequence.
ACCESSION      BI929547
VERSION      BI929547.1 GI:16243679
KEYWORDS      EST.
SOURCE      Lycopersicon esculentum (tomato)
ORGANISM      Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 800)
van der Hoeven,R.S., Bezzerides,J.L., Karamycheva,S.A., Tsai,J.,
Utterback,T., Van Aken,S., Ronning,C.M., Nierman,W., Fraser,C.M.,
Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
Generation of ESTs from tomato flower tissue, 3-8 mm buds (2001)
Unpublished
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
This clone is available through the Clemson University Genomics
Institute

```

```

FEATURES
source
Seq primer: T3.
Location/Qualifiers
1..800
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cTOB29E10"
/tissue_type="flower"
/dev_stage="3-8mm buds"
/clone_lib="tomato flower, 3 - 8 mm buds"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Cornell University; sequencing: the
Institute for Genomic Research. Flower buds and flowers
were taken from greenhouse plants (4-8 wks old, TA496).
They were immediately frozen in liquid nitrogen and then
size-separated while remaining frozen."
BASE COUNT      200 a 182 c 198 g 220 t
ORIGIN
Query Match      22.8%; Score 333.6; DB 12; Length 800;
Best Local Similarity 65.6%; Pred. No. 2.3e-84;
Matches 521; Conservative 0; Mismatches 264; Indels 9; Gaps 2;
QY      677  TCCTCTTCCAGGCCATCAACGTCACGACGCTGTACAAAGTCCAAAGTTGATAACATCT 736
Db      5  TGTCTTTCCCTGCTATTATGTTAATGACTCTGTACCAGAGCAAGTTTGACAACTTGT 64
QY      737  ACGGCTGCCGCGCACCTCCCTTATCGATGGTATCAACCGTCTTCGGATGTGATGATCGGG 796
Db      65  ACGGATGCCGCCACTCACTTCCCGATGGTCTCATGAGGCTACTGATGTTATGATTCGT 124
QY      797  GCAAGACAGCTCTGCTATGNGGTTACGGCGATGTGGCAAGGGCTGGGCTCAATCCCTCC 856
Db      125  GAAAGGTTGCCCTTGTGCTGTTATGGAGATGTGGCAAGGGATGTGCTGCTGCCATGA 184
QY      857  GTGGCCAAAGCGCTCGCGTTATCATCAGAACTGACCCCATCTGCGCTCTCCAGGCTG 916
Db      185  AACAGCTGTGTGACGCGTTATGTGACTGAGATTGATCCAACTCTGTCTCCAGGCTA 244
QY      917  CCATGGAAGGCTACCAGGTCGCGGCATCGAGGAAGTGTCTCAAGGATGTGATATCTTCG 976
Db      245  CCATGGAAGGCTCTCCAGGTTCTTCTCTTGAGGATGTTGTTCTGAGGTTGATATCTTTG 304
QY      977  TTATATGACAGAGAACTGCGATATGATCTCTGTGCAATGATGGCCCCAGATGAAGATA 1036
Db      305  TGACCACCACTGGTAAACAGGACATCATATGTTGGTTGACCACATGAGGAAGATGAAGA 364
QY      1037  AGGCTATTGTCCGTAAACATCGGCCACTTTCGATAACGAATTCATACAGATGGGCTCATGA 1096
Db      365  ATGCCATGCTTGAACATTTGGTCACTTTGACATGAAATCGACATGCTATGGTCTTGAGA 424
QY      1097  AATACCCAGGCAATCAAGCACATCCCAATCAAGCCAGAATPAGACATGTGGGAATCCCG 1156
Db      425  CCTTCCCTGTTGTAAGAGGATCACAATCAAGCCTCAAAACGACAGATGGGTTTCCCG 484
QY      1157  ATGGCCACGCTATCTCCCTCTTCTT--GCTGAGGCGCGCTTCTTAACCTTGGTGGCTA 1213
Db      485  ACACCAACAGTGGCATCATTTGTGTGGCCGAGGTCGTCTCATGAACTTGGGATGTGCCA 544
QY      1214  CAGTCAACCATCTTTCGTTATGTCAATGTCAATTCACAAACACAGACACTGGCTCAGCTCG 1273
Db      545  CAGGACACCCAGTTTGTGATGTTCTTCTTCTTCACTAACCAAGTCAATGGCCCACTCG 604
QY      1274  ACTTCT-----ACGAAGAAGAGGAATCTCGAGAAGAGGTTTACACACTTCCGAAGC 1327
Db      605  AGTTGTGAATGAGAAGAGCAGTGGTAAATACGAGAAAGAGATATACGTTTTCGCCAAGC 664
QY      1328  ATCTCGATGAAGAAGTCCGCTCGCTCCACTCGATCTCGATCTCCAGTCCACTTCAAGAC 1387
Db      665  ACCTTGACAGAGAGTTGCTGCCCTTCACTCTTGGAAGCTCGGAGCCAAACTTACCAAC 724
QY      1388  TTACACAGAAGCAGGCTGACTACATCAATCAACGTTTCCAGTTGAGGGTCTTACAGATCTGATG 1447

```

```

Db      725 TTTACAGGATCAGCTGACTACATTAGCTACCAAGTTGAGGGTCTCTTTTACAGCTGCTC 784
QY      1448 CTTACCGTTATTAA 1461
Db      785 ACTACAGGACTGA 798

RESULT 12
CB682756
LOCUS   CB682756
DEFINITION Oryza sativa (japonica cultivar-group) cDNA
ACCESSION OSUNEF10G13.f OSUNEF Oryza sativa (japonica cultivar-group) cDNA
VERSION   CB682756
KEYWORDS  CB682756.1 GI:29686481
SOURCE   EST.
ORGANISM Oryza sativa (japonica cultivar-group)
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
          Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 774)
AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
          Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
TITLE     Large-scale identification of ESTs involved in the interaction
          between rice and Magnaporthe grisea
JOURNAL   Unpublished
COMMENT   Contact: Rod Wing
          Arizona Genomics Institute
          University of Arizona
          Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
          85721-0088, USA
          Tel: 520 626 3967
          Fax: 520 621 9288
          Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 10 row: G column: 13
Seq primer: gta aaa cga cgg cca gtc.

FEATURES
source
1..774
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/db_xref="taxon:39947"
/clone="OSUNEF10G13"
/tissue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSUNEF"
/notes="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
XhoI; Uninfected Control"
BASE COUNT 168 a 220 c 205 g 180 t
ORIGIN
Query Match 22.8%; Score 333; DB 14; Length 774;
Best Local Similarity 67.9%; Fred. No. 3.4e-84;
Matches 497; Conservative 0; Mismatches 226; Indels 9; Gaps 2;

QY      612 CGGTGTTTCGAGAGACAACACAGGTGTCCACCGCTCTACAGCTCGAGAAGGAGG 671
Db      42 CGGAGTCTCGAGGAGACCACCACCGGTGTCAAGAGGCTTACCAAGATGAGGAGCCGG 101
QY      672 CAATCTCTTCCAGGCATCAACGTCACGACGCTGTACAAAGTCCAAAGTTCGATAA 731
Db      102 CGCCCTCTCTTCCCGCCATCAACGTCACGACTCCGTCACCAAGACGAAGTTGACAA 161
QY      732 CATCTAGGCTGCGCCCACTCCCTATCTGATGATGATATCACCGTGTCCGATGATGAT 791
Db      162 CCTGATGTTGCGGCCACTCTCTCCCTGATGTTCTCATGAGGCTACCGATGATGAT 221
QY      792 CGGGGCAAGACGCTCTGTCATGGTTACGGCGATGTCGGCAGGCGCTCAATC 851

```

```

Db      222 CGCTGGCAAGGTTCCGCTGCTCGCGTTATGCTGATGTTGGCAAGGCTGTGCTGCTC 281
QY      852 CTTCGCTGGCCAGGCGCTCGGTTATCATCACAGAACTCGACCCCAATCTGGCTCTCCCA 911
Db      282 TCTCAAGCAGGCTGTGCTCCGCTGCTCATTTGTACTGAGATTGACCCCATCTGTGCCCTCCA 341
QY      912 GGCTGCCATGGAAGGCTACCCAGGTCCGCGCATCGAGAAAGTCGTCAAGGATGTCGATAT 971
Db      342 GGCCCTTATGAGAGGCTCTCCAGGTCTCCAGTCTCTGAGGATGTTGCTCGAGGCTGACAT 401
QY      972 CTTCGTTATGTCACAGAAACTCGATATCATCTCTGTTGACATGATGGCCAGATGAA 1031
Db      402 CTTTGTGACCACTGTCGCAACAGGACATCATATGTTGACCATGAGGAAGATGAA 461
QY      1032 GGATAAGGCTATTGTCGTTAACTGGCCACTTCGATAACGAAATTCATACAGATGGCT 1091
Db      462 GAACATATGCCATGCTTTCACATATGCTGCTTTCACATGATGATGCTCGGCT 521
QY      1092 CATGAAATACCCAGGCTATCAAGCACAATCCCAATCAAGCCAGATACGATGTGGAAAT 1151
Db      522 TGAGACCTACCTGCTGCTCAAGGCGATCACCATCAAGCCTCAGACCGACCGCTGCTT 581
QY      1152 CCAGATGGCAGCCT---ATCCTCTTCTTCTGAGGCGCGCTTCTTAACCTTGGCTG 1208
Db      582 CCTGAGACCAACTGGCATCTTCTTCTGCTGAGGGTCTCTCATGAACCTTGGGTG 641
QY      1209 CGCTACAGTCAACCCATCTTTCTGTTATGTCATATGTCATCAACAAACAGACACTCGTCA 1268
Db      642 CGTACTGGCCACCCAGTTTGTGCTGCTGCTCTCACTTCACTAACAGGTCAATGCTCA 701
QY      1269 GCTGACCTCT-----ACGAAAGAGAGGAAATCTCGAAGAGAGTTTACACTTCC 1322
Db      702 GCTTGAGCTGTGGAAGGAGAGACCTGGCAAGTACGAGAGGTTGCTACGTTCTTCC 761
QY      1323 GAAGCATCTCGA 1334
Db      762 NCAGCACTCGA 773

RESULT 13
BO744296
LOCUS   BO744296
DEFINITION WHE4113_H10_O192S Wheat salt-stressed root cDNA library Triticum aestivum cDNA clone WHE4113_H10_O19, mRNA sequence.
ACCESSION BO744296
VERSION   BO744296.1 GI:21891083
KEYWORDS  EST.
SOURCE    Triticum aestivum (bread wheat)
ORGANISM  Triticum aestivum
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
          ; Triticeae; Triticum.
          1 (bases 1 to 748)
          Anderson,O.D., Akhunov,E., Chao,S., Crossman,C., Deal,K., Dvorak,J.,
          Lazo,G.R., Phan,J., Raush,C.J., Wilson,C. and Woo,J.
          The structure and function of the expressed portion of the wheat
          genomes - Salt-stressed root cDNA library
          Contact: Olin Anderson
          US Department of Agriculture, Agriculture Research Service, Pacific
          West Area, Western Regional Research Center
          800 Buchanan Street, Albany, CA 94710, USA
          Tel: 5105959773
          Fax: 5105959818
          Email: candersn@pw.usda.gov
          Sequences have been trimmed to remove vector sequence and low
          quality sequence with phred score less than 20
          Seq primer: SK primer.
          Location/Qualifiers
            1..748
              /organism="Triticum aestivum"
              /mol_type="mRNA"

FEATURES
source

```

/cultivar="Chinese Spring"  
/db xref="taxon:4565"  
/clone="WHB4113 H10\_O19"  
/tissue\_type="Roots"  
/dev\_stage="Full tillering"  
/lab\_host="E. coli SOLR"  
/clone\_lib="Wheat salt-stressed root cDNA library"  
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid  
pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Hydroponic  
plants grown to full tillering stage were treated with 150  
mM NaCl for either 12 hours or 7 days. Root tissues of the  
plants subjected to both types of treatment were collected  
separately at University of California, Davis (E. Ahumov  
and K. Deal in J. Dvorak's Lab). Total RNA was prepared  
separately from the two samples (12h and 7day treatments),  
and equal amount of RNA was then pooled. PolyA RNA was  
purified from the pooled RNA, a cDNA library was made, and  
the cDNA clones were in vivo excised to give pBluescript  
SK(-) phagemids in J. Dvorak's lab (E. Ahumov, J. Dvorak)  
at the University of California, Davis. Colony plating,  
plasmid DNA preparations and DNA sequencing were performed  
in the OD Anderson lab (all other authors)."

BASE COUNT 165 a 214 c 203 g 166 t  
ORIGIN

Query Match 22.8%; Score 332.8; DB 13; Length 748;  
Best Local Similarity 69.4%; Pred. No. 3.8e-84;  
Matches 470; Conservative 0; Mismatches 207; Indels 3; Gaps 1;

QY 612 CGGTGTTTCGAGAGACAAACAGGTGCCACGGCTCTACCGCTCGAGTCGAGAGGAGG 671  
Db 53 CGGCGTCTCCGAGGAGACCACTACCGGCTCAAGAGCTCTACAGATCGAGGAGTCCGG 112  
QY 672 CAATCTCTCTCCAGCCCATCAAGTCAACGAGCTGTTCACAAAGTCCAAAGTTCGATAA 731  
Db 113 CACCCTCTCTCCCGCCCATCAAGTCAAGACTCGTTCACCAAGAGCAAGTTTGACAA 172  
QY 732 CATCTACGGTGGCGGCATCCCTTATCGATGATATCAACCGTGTTCGATGTCATGAT 791  
Db 173 CTTTACGGTGTGGCGGCATCCCTCTGATGCTTATGAGGCTCTGATGATGATGAT 232  
QY 792 CGCGGCAAGACAGCTCTGTCATGGTTACGGGATGTCGCAAGGCTGCGTCAATC 851  
Db 233 CGCGGCAAGGTTGCTGTGCTGCGGTATGATGATGTCGCAAGGCTGTCGCGC 292  
QY 852 CTTCCGTGGCAAGGCGCTCGCGTTATCATCAAGACTCGACCCAAATCTCGCTCTCCA 911  
Db 293 ACTCAAGCAGCTGTGTGCCGTGTGATCGTCAAGAGATTGACCCCATCTGTGCCCTTCA 352  
QY 912 GCGTCCATGGAAGGCTACAGGTCCGCGCATCGAGGAATCGTCAAGATGTCGATAT 971  
Db 353 GCGCTGATGAGGGGATCCAGATCCTCACTTTGGAGAGCTGTGCTCTGATGTCGAT 412  
QY 972 CTTGTTACATGACAGGAATCGGATATCATCTCTGTTGATCATGATGGCCAGATGAA 1031  
Db 413 CTTGTTGACCACTGGAACAGGACATCATCATGTTGACCATGAGGAAGTAA 472  
QY 1032 GATTAAGCTATTGTGTTAATCATCGGCCTTCGATACGAAATTTGATACAGTGGCCT 1091  
Db 473 GAACAATGCCATTGTCTGAACATTGGTCACTTTGACACGAGATCGACATGAACGGCCT 532  
QY 1092 CATGAATACCCGGCATCAGCATCCCATCAAGCCAGATACGATCTGGGAAT 1151  
Db 533 CGAGACCTTACCCGTGGTCAAGGGCATCACCATCAAGCCTCAGACCGAGCTGGGCTT 592  
QY 1152 CCCAGATGGCCAGCT...ATCCTCTCTTCTGCTGAGGGCGCGCTTCTTAACTTGGCTG 1208  
Db 593 CCCCGAGACCACTGGGCTCATTTGCTGTGAGGGCGCTCTGATGAACCTTGGCTG 652  
QY 1209 CGCTACAGGTACCCCATCTTTTGTGTTATGTAATGTAATTCATAAACCAGCACTCGGTCA 1268  
Db 653 TGCCACTGGCCACCCAGCTTTGTCTATGTCCTGCTCATTCCTACTAACCAAGGTCATGCTCA 712

QY 1269 GCTCGACCTCTACGAAAAGA 1288  
Db 713 GCTGAGTTGTGAAGGAGA 732  
RESULT 14  
AL163774  
LOCUS AL163774 977 bp mRNA linear EST 04-APR-2001  
DEFINITION AL163774 Scherffelia dubia pSPORT1 Scherffelia dubia cDNA clone 418  
similar to S-adenosyl homocysteinase, mRNA sequence.  
ACCESSION AL163774  
VERSION AL163774.1 GI:13547742  
KEYWORDS EST.  
SOURCE Scherffelia dubia  
ORGANISM Scherffelia dubia  
REFERENCE 1 (bases 1 to 977)  
AUTHORS Becker, B., Feja, N. and Melkonian, M.  
TITLE Analysis of expressed sequence tags (ESTs) from the scaly green  
flagellate Scherffelia dubia Pascher emend. Melkonian et Preisig  
JOURNAL Protist 152 (2), 139-147 (2001)  
MEDLINE 21428165  
PUBMED 11545437  
COMMENT Contact: Becker B  
Botanisches Institut  
Universitaet zu Koeln  
Gyrhofstr. 15, 50931 Koeln, Germany.  
FEATURES  
Location/Qualifiers  
1..977  
/organism="Scherffelia dubia"  
/mol\_type="mRNA"  
/strain="SAG 40.89"  
/db\_xref="taxon:3190"  
/clone="418"  
/clone\_lib="Scherffelia dubia pSPORT1"  
BASE COUNT 216 a 330 c 276 g 154 t 1 others  
ORIGIN  
Query Match 22.6%; Score 330.2; DB 9; Length 977;  
Best Local Similarity 64.5%; Pred. No. 2.5e-83;  
Matches 522; Conservative 1; Mismatches 284; Indels 2; Gaps 2;  
QY 481 TTCGAACAGCCGGTCTGTTCCAGAGCCAAAGAGCTGCAACCTCGACACAGTTGCT 540  
Db 93 TTGCCAAGAGCGCCACCATCCCAACCCCGAGTCCACCGAACCCAGAGTTCAAGATC 152  
QY 541 GTTCTTCTACTCAAGCAGGTCTTCAACCAAGACAAGAACCTGCGACACAGTTGCT 600  
Db 153 GTGTCACCTCATCAAGAGGGCTGCAAGAGACCCCGAAGTACCAGATGGCC 212  
QY 601 CGCGCATGAAGGTGTTTCGAAGAGACAACACAGGTGTCACCGCTCTACCACTC 660  
Db 213 AAGCGCTGTGAGGGTCTCCGAGGAGACACACCGCGGTGAAGC-STTGTACGAGTG 271  
QY 661 GAGAGAGAGGCAAACTCTCTCCAGCCCATCAAGTCAAGCGCTGTGTACAAAGTCC 720  
Db 272 CAGACAAAGGCTGCGTGTCTTCCCGCATCAAGTCAAGCTCGGTGACAGATGCC 331  
QY 721 AAGTTCGATAAATCTAAGGGTGGCGCACTCCCTTATCGATGGTATCAACCGGTCTTC 780  
Db 332 AAGTTCGACAAAGTGTACGGCTGGCGCACTCTCCCGCGGCACTATGCGGGCCATC 391  
QY 781 GATCTCATGATCGCGGCAAGAGCTCTCTGTCATGGGTTACGGCGATGTCGCAAGGSC 840  
Db 392 GACGCTATGTCGAGGCAAGACCGCTCTTTCATCGTGGCTTCGGTGAAGTGGGCAAGGSC 451  
QY 841 TGCGCTCAATCCCTCCGTGGCCAGAGGCTCGGTTATCATCAGAACTCGACCAATC 900  
Db 452 TGCTGCGCCGCATGAAGGCGCGCGCGGCACCATCGTCTCGAGATCGACCCCATC 511  
QY 901 TGCGCTCTCCAGGCTGCCATGAAGGCTTACAGGTCCGCGGCATCGAGAGTCTGTCAG 960



```

Db      512 TCGCGCTCGAGCGCTGATGAGCGGTACCGAGTGTGCGGTGAGGAGTGGCTCAGG 571
Qy      961 GATGTCGATATCTTCGTATCATGACAGGAACTCGGATATCATCTCTGTTGACATGATG 1020
Db      572 TACGCCGACATCTTCATCACCACCGGCAACAGGATATCATCATGCTGGAGCACAATG 631
Qy      1021 GCCAGATGAGGATGAGCTATGTCGTTACATCGGCCACTTCGATGAAAGAAATGAT 1080
Db      632 GCCAAGATGAGCAACAGCCATCGTGGCAACATCGGCCACTTCGACAGAGATCGAC 691
Qy      1081 ACAGATGCGCTTCATGAAATACCCAGGATCAAGCACATCCCAATCAAGCCAGAAATACGAC 1140
Db      692 ATGGCGGCGCTCTTCAAGTACTCCGGCATCAGCGCCAGACATCAAGCCGACGTGGAC 751
Qy      1141 ATGTGGGAATTCAGATGCGCACGCTATCTCTCTTGTGAGGGCGGCTTCTTTAAC 1200
Db      752 AAGTTCACTTCCG-GACGGCACTCCATCATCATCTCGGCCGAGGCGCGCTCTTAAC 810
Qy      1201 CTTGGCTCGCTACAGCTCAGCTCACCATCTTTCGTTATGTCATGTCATTCACAAACGACA 1260
Db      811 CTGGGCTCGGCCACCGGACACCCCTCTTCTGTCATGCTCTTCTTCTTTCACAAACGAGTC 870
Qy      1261 CTCGCTCAGCTCGACCTCTACGAAAGAG 1289
Db      871 ATCGCGCACTGGGAGCTGTGGAACGAGCG 899

RESULT 15
CB618580
LOCUS      779 bp mRNA linear EST 08-APR-2003
DEFINITION Oryza sativa (indica cultivar-group) cDNA
clone OSIIeA02J06 5', mRNA sequence.
CB618580
ACCESSION CB618580.1 GI:29613567
VERSION    1
KEYWORDS   Oryza sativa (indica cultivar-group)
SOURCE     Oryza sativa (indica cultivar-group)
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Shrharoidae; Oryzaceae; Oryza.
REFERENCE  1 (bases 1 to 779)
AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
TITLE     Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
JOURNAL   Unpublished
COMMENT   Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat sac cat g
Plate: 02 row: J column: 06
Seq primer: gta aaa cga cgg cca gtc.
Location/Qualifiers
1..779
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="IR36"
/db_xref="taxon:39946"
/clone="OSIIeA02J06"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSIIeA"
/note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
XhoI; Lesion Mimic SPL ii"
BASE COUNT 166 a 233 c 204 g 176 t

```

```

ORIGIN
Query Match      22.5%; Score 329.4; DB 14; Length 779;
Best Local Similarity 65.1%; Pred. No. 3.7e-83;
Matches 502; Conservative 0; Mismatches 266; Indels 3; Gaps 1;
Qy      499 GTTCAGAGGCAACAGAGCTGACAACTCGAATACCGCTGGTCTCTGCTACACTCAAG 558
Db      9 GTCCCGACCGCGAGTCCACGACACCGCGAGTTCAAGATCGTCTCACCATCATCCGG 68
Qy      559 CAGCTCTTCAACAGACAAAGAACCACTGGCACACAGTTTGTCTGCCGCGCATGAACGGTGT 618
Db      69 GACGGCTCAAGTCCGACCCAGCAAGTACCGCAAGATGAAGAGAGGCTCTGTCGAGTC 128
Qy      619 TCCGAAGAGACAAACAGAGTGTCCACCGCTCTACACAGCTCGAGAGAGGGGCAACTC 678
Db      129 TCCGAGGAGACCAACCAACCGGTGTCAAGAGGCTCTACCAAGATGCAGGAGACCGCGGCCCTC 188
Qy      679 CTCTTCCAGCGCATCAAGCTCAAGCGCTGTACAAAGTCCAAAGTTCGAATACATCTAC 738
Db      189 CTCCTTCCCGCCATCAAGCTCAAGCACTCCGCTCACCAGAGCAAGTTTGAACAACCTGTAT 248
Qy      739 GCCTGCCGCCACTCCCTTATCGATGATGATCAACCGTGTCTTCGATGTCTCATGATCGCGGC 798
Db      249 GGTTCGCCGCCACTCTCTCCCTGATGCTCTCATGAGGCTTACCGATGTTTATGATCGCTGC 308
Qy      799 AAGACAGCTCTCGTCATGGGTTACGGCGATGTCGGCAAGGGCTCGCTCAATCCCTCCGT 858
Db      309 AAGGTTCCGCTGCTGCGGTTATGTTGATGTTGGCAAGGGCTGTGCTGTCTCTCAAG 368
Qy      859 GGCACAGCGCTCGCTTATCATCAGAGAACTCGACCAATCTCGCTCTCCAGGCTGCC 918
Db      369 CAGGCTGTGCGGCTCATCGTACTGAGATTGACCCCATCTGTCCCTCCAGGCCCTT 428
Qy      919 ATGGAAGGCTTACCAAGTCCCGCGCATCGAGAAAGTCTGTCAGGATGTGATATCTTCGTT 978
Db      429 ATGGAGGCTCTCCAGTCTCTCACTTGGAGGATGTTGTCTCGAGGCTGACATCTTTGTG 488
Qy      979 ACATGCAAGAGAACTGGGATATCATCTCTGTTGACATGATGCCCGCAGATGAAGATAAG 1038
Db      489 ACCACCACTGGCAACAAAGGACATCATTAATGTTGACCAATGAGGAAGATGAAGACAAT 548
Qy      1039 GCTATTGTGCGTAACATCGGCCACTTGGATAAGAAATTGATACAGATGGCTCATCAAA 1098
Db      549 GCCATCGTTTGCAACATTTGGTCACTTTTGAACATGAGATTGACATGCTCGGCTTGAGACC 608
Qy      1099 TACCAGGCTCAAGCAATCCCAATCAAGCCAGAAATGACATGTGGAAATCCCAAGAT 1158
Db      609 TACGCTGTGTCAAGCGCATCAACCATCAAGCTCAGACCGGCGCTGCGTCTTCCCTGAG 668
Qy      1159 GGCCACGCT---ATCCTCCTTCTGCTGAGGCGCGGCTTCTTAACCTTGGCTGGCTACA 1215
Db      669 ACCAACACTCGCATCATTTGCTCTGCTGAGGCGGTCTCATGAACCTTGGGTGGCTACT 728
Qy      1216 GGTCAACCATCTTTCTGTTATGTCATATGTCATTTCACAAACGACACTCGCT 1266
Db      729 GGCAACCCAGTTTCTCATGTCTCTGCTCATTAACAGGTCAATTGCT 779

Search completed: October 17, 2003, 22:53:46
Job time : 3335 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: October 15, 2003, 14:33:49 ; Search time 99.5 Seconds  
(without alignments)  
4661.297 Million cell updates/sec

Title: US-09-759-990A-1

Perfect score: 2737  
Sequence: 1 atgcttgcaatcacctac.....ctgatgcttaccgtattataa 1461

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 2215726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -US0975990/runat 15102003 115151 2344/app query.fasta\_1.1607  
-Q=/cgn2.1/USPTO spool/US0975990 -SUPFIX=rag -MINMATCH=0.1 -LOOPCL=0  
-DB=A Geneseq 19Jun03 -QWIT=fastan -SUPFIX=rag -MINMATCH=0.1 -TRANS=human40.cdi  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFM=prco -NORM=ext -HSPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0975990 @CEN 1 1 137 @runat 15102003 115151 2344 -NCPU=6 -ICPU=3  
-NO.MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPEBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A Geneseq 19Jun03.\*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1580	57.7	469	24	ABU11410 Protein encoded by
2	1550.5	56.6	495	21	AA978783 M. tuberculosis an
3	1537	56.2	485	21	AAG40086 Arabidopsis thalia
4	1537	56.2	508	21	AAG40085 Arabidopsis thalia
5	1526	55.8	486	13	AAR26500 Prod. of Nicotiana
6	1523	55.6	485	21	AAG34060 Zea mays protein f
7	1523	55.6	510	21	AAG34059 Zea mays protein f
8	1519	55.5	467	21	AAG40087 Arabidopsis thalia
9	1512	55.2	485	21	AAG41407 Arabidopsis thalia
10	1512	55.2	497	21	AAG41406 Arabidopsis thalia
11	1496	54.7	485	17	AAW01459 Arabidopsis S-aden
12	1494	54.6	467	21	AAG41408 Arabidopsis thalia
13	1490	54.4	485	17	AAW01458 Asparagus S-adenos
14	1473	53.8	450	21	AAG34061 Zea mays protein f
15	1457	53.2	467	21	AAG22957 Arabidopsis thalia
16	1457	53.2	489	21	AAG22956 Arabidopsis thalia
17	1454	53.1	474	22	AAG90582 C. glutamicum prote
18	1454	53.1	478	22	AAU71911 C. glutamicum meta
19	1454	53.1	478	22	AAU79743 Corynebacterium gl
20	1454	53.1	498	23	AAU97842 Corynebacterium gl
21	1439	52.6	449	21	AAG22958 Arabidopsis thalia
22	1318	48.2	432	22	AAU71913 C. glutamicum meta
23	1318	48.2	432	22	AAU79745 Corynebacterium gl
24	1269.5	46.4	433	23	AAE25380 Human NZMS-4 prote
25	1266.5	46.3	432	22	ABU70895 Drosophila melanog
26	1260	46.0	457	21	ABU43614 Human cancer assoc
27	1081.5	39.5	504	22	ABU71710 Drosophila melanog
28	1057	38.6	521	22	ABU38663 Drosophila melanog
29	1034	37.8	564	23	ABG32371 Novel human enzyme
30	1033	37.7	500	20	AAW90061 Human SAHH protein
31	1032	37.7	614	19	AAW56097 Amino acid sequenc
32	958	35.0	464	21	AAU77950 A. thaliana enviro
33	735.5	26.9	422	22	AAU96504 Putative P. abyssi
34	687	25.1	455	22	ABG12029 Novel human diagno
35	606	22.1	395	23	AAE25390 Human NZMS-14 prot
36	551.5	20.1	170	23	AAU49455 Rat S-adenosylhomo
C 37	541	20.7	236	22	AAG90583 C. glutamicum prote
C 38	426.5	16.3	240	22	AAG90584 C. glutamicum prote
39	413.5	15.1	500	23	ABP65910 Bifidobacterium lo
40	281	10.3	81	21	AAG01948 Human secreted pro
41	281	10.3	97	16	AAU80298 Human SAHH. Homo
42	277	10.1	95	16	AAU80297 SAHH-GSE-encoded p
43	266.5	9.7	312	22	ABG12030 Novel human diagno
44	249	9.1	66	23	ABP42348 Human ovarian anti
45	243	8.9	135	22	AAO10323 Human polypeptide

# ALIGNMENTS

RESULT 1  
ABU11410  
ID ABU11410 standard; Protein; 469 AA.

XX AC ABU11410;

XX DT 11-FEB-2003 (first entry)

DE Protein encoded by S. atroolivaceus leinamycin gene cluster ORF+8.

XX Leinamycin biosynthesis gene cluster; Lmm; open reading frame; ORF;

KW anti-tumour antibiotic; broad spectrum antimicrobial activity;

KW Gram-positive; Gram-negative bacteria; chemical modification;

KW metabolite; apo-carrier protein; holo-carrier protein; tumour;

KW polyketide; hybrid polypeptide/polyketide metabolite; Lmm production;

XX Streptomyces atroolivaceus.

OS

XX WO200277179-A2.  
 PN  
 XX  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 22-MAR-2002; 2002WO-US089937.  
 XX  
 XX 26-MAR-2001; 2001US-278933P.  
 PR  
 XX (REGC) UNIV CALIFORNIA.  
 PA (KYOW) KYOWA HAKKO KOGYO KK.  
 XX  
 XX  
 PI Shen B, Cheng Y, Tang G;  
 DR WPI; 2003-018907/01.  
 XX N-PSDB; ABX34289.  
 PT Novel gene cluster responsible for synthesis of leinamycin in  
 PT Streptomyces atroolivaceus useful for making various peptide and/or  
 PT polyketide, and/or hybrid polypeptide/polyketide metabolites -  
 XX  
 PS Claim 13; Page 155; 185pp; English.  
 CC The present invention relates to the isolation of the Streptomyces  
 CC atroolivaceus leinamycin (lmm) biosynthesis gene cluster containing  
 CC 71 open reading frames (ORFs) (ORFs -35 through -1, ORFs lmmA through  
 CC lmmZ, and ORFs +1 through +9). Leinamycin is a novel anti-tumour  
 CC antibiotic produced by several Streptomyces species. It exhibits  
 CC broad spectrum antimicrobial activity against Gram-positive and  
 CC Gram-negative bacteria, but not against fungi. The polypeptides encoded  
 CC by the lmm biosynthesis gene cluster ORFs are useful for chemically  
 CC modifying a molecule in a host cell. The host cell is a bacterium or  
 CC eukaryotic cell, including a mammalian, yeast, plant, fungal, or insect  
 CC cell. The molecule is an endogenous metabolite produced by the host  
 CC cell or exogenously supplied metabolite, or an amino acid, and the  
 CC polypeptides encoded by the lmm gene cluster are useful for converting  
 CC an apo-carrier protein to a holo-carrier protein. lmm shows potent  
 CC antitumour activity in tumour models in vivo. The lmm gene cluster  
 CC modules and/or catalytic domains are useful for making various peptide  
 CC and/or polyketide, and/or hybrid polypeptide/polyketide metabolites.  
 CC The proteins encoded by the ORFs are useful alone, or in combination  
 CC with other active domains to modify various target substrates. The  
 CC lmm gene cluster is useful to upregulate endogenous lmm production to  
 CC permit lmm production in cells and/or to make various modified lmm.  
 CC lmm, its analogue, or other polyketide, peptide or hybrid  
 CC polyketide/peptide metabolites are useful as therapeutic agents, to  
 CC treat a number of disorders, depending upon the type of metabolites.  
 CC ABU11341-ABU11411 represent the proteins encoded by ORFs of the  
 CC S. atroolivaceus leinamycin biosynthesis gene cluster.  
 XX  
 SQ Sequence 469 AA;  
 Alignment Scores:  
 Pred. No.: 3.9e-131 Length: 469  
 Score: 1580.00 Matches: 307  
 Percent Similarity: 77.26% Conservative: 60  
 Best Local Similarity: 64.63% Mismatches: 92  
 Query Match: 57.73% Indels: 16  
 DB: 24 Gaps: 3  
 US-09-759-990A-1 (1-1461) x ABU11410 (1-469)  
 QY 34 GAGTACAGAAATCCGACATCAACCTCCATGTTCTCGCCGCTAGGAACCTTACCCCTGCT 93  
 DB 11 ASPHLYSVALAASPLeuSerLeuAlaPheGlyArgLysGluIleThrLeuAla 30  
 QY 94 GAGAGGAAATCCGAGGCTTATGGTTCTCGTGGCCGCTTATCCGCTTCTAAGCCATTG 153  
 DB 31 GluHisGluMetProGlyLeuMetSerIleArgGluGluTyralaAlaAlaGlnProLeu 50  
 QY 154 AAGGGTGTCAATCTCGTTCCCTCCATGACATGACAGTCCAGACCGCTCTCATCAG 213

Db 51 AlaGlyAlaArgIleThrGlySerLeuHisMetThrValGlnThrAlaValLeuIleGlu 70  
 QY 214 ACATCACAAGCTCTTGGTGCTGATGTCAGATGGCTTCCTGCAACATCTCTCTACACAA 273  
 Db 71 ThrLeuValAlaLeuGlyAlaAspValArgTrpValSerCysAsnIleTyrSerThrGln 90  
 QY 274 GATACAGCCGCTGCTGCTATCGTTGTCGGCCCAACAGGCACACAGGAGAACAGCCGCGT 333  
 Db 91 AspHisAlaAlaAlaIleAlaAla-----AlaGly 101  
 QY 334 ATCCAGCTCTTCGCTGGAGGCGAAACACTCCAGAAATCTGGAGAACACATACCCG 393  
 Db 102 IleProValPheAlaTrpGlyGluThrLeuGluGluTyrTrpCysThrGluGln 121  
 QY 394 CTTCTCATGCGCAGATGGTCAAGGCCACACAGCGTTGTGTGATGTGTGTGTGTGTGTGT 453  
 Db 122 AlaLeuThrTrpProGlyHisThrGlyProAsnMetIleLeuAspAspGlyGlyAspAla 141  
 QY 454 ACATCTCTCATCTCCAGGGCTTCGAATTCGAACAGCGGTGCTGTTCAGAGCCACA 513  
 Db 142 ThrLeuLeuValHisGlyGlyValGluTyrArgLysThrGlyIleLeuPro----- 158  
 QY 514 GAAGCTCACAACTCGAATACCGCTGGCTTCTGCTTACACTCAAGCAGGCTTCAACCAA 573  
 Db 159 GluAlaGluAsnGluGlu-----LeuAlaValValArgAlaLeuLeuAspArg 174  
 QY 574 GACAGAACCATCTGGCACAACAGTGTGTCGCGCATGAGCGGTGTTCCGAGAGACA 633  
 Db 175 SerGlyLeuAspTrpThrAlaMetSerSerGluIleArgGlyValThrGluGluThrThr 194  
 QY 634 ACAGGTCTCCACGCTCTTACCAGCTCGAGAGGGGCAAACTCTCTCCAGCGCATC 693  
 Db 195 ThrGlyValHisArgLeuTyGluMetHisArgAspGlyThrLeuLeuPheProAlaIle 214  
 QY 694 AACCTCAACGACGCTGTTCACAAAGTCCAAAGTCCGATACATCTACGCTCCGCCACTCC 753  
 Db 215 AsnValAsnAspAlaValThrLysSerLysPheAspAsnLysTyrGlyCysArgHisSer 234  
 QY 754 CTTATCATGATGATCAACCGTCTCCGATGTCATGATCGCGGCAACAGCTCTCGTC 813  
 Db 235 LeuIleAspGlyIleAsnArgAlaThrAspValLeuIleGlyLysThrAlaValAlaVal 254  
 QY 814 ATGGTTTACCGCCATGTCGCAAGGCTCGCTCAATCCTCGTGGCCAAAGCGCTCGC 873  
 Db 255 CysGlyTyrGlyAspValGlyLysGlyCysAlaGluSerLeuArgGlyGlnGlyAlaArg 274  
 QY 874 GTTATCATCAGAACTCGACCCAAATGTCGCTCTCCAGCTGCCATGGAAGGCTACCAG 933  
 Db 275 ValIleIleThrGluIleAspProIleCysAlaLeuGlnAlaAlaMetAspGlyTyrGln 294  
 QY 934 GTCCGCCGATCGAGGAGTCTGCAAGGATGTCGATATCTCGTTACATGACAGGAAAC 993  
 Db 295 ValAlaThrLeuAspGluValValAspLysAlaAspIlePheValThrThrGlyAsn 314  
 QY 994 TCCGATATCATCTCTGTTGACATGATGCCCAAGATGAAGCTATGTCGGTATC 1053  
 Db 315 LysAspIleIleMetAlaSerAspMetAlaLysMetLysHisGlnAlaIleValGlyAsn 334  
 QY 1054 ATCGGCCACTTCATACGAAATTCATACAGATGGCTCATGAAATACCCAGGATCAG 1113  
 Db 335 IleGlyHisPheAspAsnGluIleAspMetAlaGlyLeuAlaGlnIleProGlyIleVal 354  
 QY 1114 CACATCCCAATCAAGCAGATACGATGTGGGAATTCACAGATGGCCAGCTATCCTC 1173  
 Db 355 LysAspGluValLysProGlnValHisThrTrpLysPheProAspGlyLysValLeuIle 374  
 QY 1174 CTTCTGCTGAGGCGCGCTTCTTAACCTTGGCTGCGCTACAGGTACCCATCTTCGTT 1233  
 Db 375 ValLeuSerGluGlyArgLeuLeuAsnLeuGlyAsnAlaThrGlyHisProSerPheVal 394  
 QY 1234 ATGTCATGTCATTCACAAACAGACACTCGCTCAGCTCGACTCTCAGAAAGAGAGGA 1293  
 Db 395 MetSerAsnSerPheAlaAspGlnThrLeuAlaGlnIleGluLeuPheThrLysProAsp 414

QY 1294 AATCTCGAAGAGGTTTACACACTCCGAAGCATCTCGATCAAGAGTGGCTCGCTC 1353  
 Db 415 GluTyrProThrAspValTyrValLeuProLysHisLeuAspGluLysValAlaArgLeu 434  
 QY 1354 CACCTCGGATCTTCGATGTCACCTTACAAAGCTTACACAGAGCGGCTGATCATC 1413  
 Db 435 HisLeuAspAlaLeuGlyValLysLeuThrThrLeuArgProGluGlnAlaSerTyrIle 454  
 QY 1414 AACGTTCCAGTTGAGGCTCTTACAAAGTCTGATGCTTACCGTTAT 1458  
 Db 455 GlyValGluValAspGlyProTyrLysProAspHisTyrArgTyr 469

## RESULT 2

RAY87873  
 ID AAY87873 standard; Protein; 495 AA.  
 AC AAY87873;  
 XX  
 DT 06-OCT-2000 (first entry)  
 DE M. tuberculosis antigen TB54 protein.  
 XX  
 KW Tuberculosis; TB; antigen; vaccine; diagnosis; somatic; tuberculostatic;  
 KW infection; interferon-gamma; IFN-gamma; protective immunity; therapy;  
 KW delayed type hypersensitivity response; TB54.  
 XX  
 OS Mycobacterium tuberculosis.  
 PN W0200021983-A2.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 08-OCT-1999; 99WO-DK00538.  
 XX  
 PR 08-OCT-1998; 98DK-0001281.  
 PR 21-JAN-1999; 99US-0116673.  
 XX  
 PA (STAT-) STATENS SERUM INST.  
 XX  
 PI Andersen P, Welling K, Hansen CV, Florio W, Okkels LMM;  
 PI Skjot RV, Rosenkrands I;  
 XX  
 DR WPI; 2000-317931/27.  
 DR N-PSDB; AAA39367.  
 XX

PT Novel polypeptide of somatic protein extract useful as vaccine against  
 PT virulent Mycobacterium infection, isolated from cell wall, cell  
 PT membrane and cytosol -  
 XX

PS Claim 1; Page 96-97; 126pp; English.

XX This invention describes a novel polypeptide (PP) of somatic proteins  
 CC extract (I) which have tuberculostatic activity. (I) or their subsequence  
 CC has at least one of the following properties: (a) the PP induces an in  
 CC vitro recall response, or an in vitro response, during primary infection  
 CC with virulent Mycobacterium, determined by a release of interferon  
 CC (IFN-gamma). (b) PP induces a protective immunity, determined by  
 CC vaccinating an animal with PP and an adjuvant, three times at two weeks  
 CC intervals, (c) PP induces an in vitro response, or in vitro recall  
 CC response, determined by release of IFN-gamma of at least 1000 pg/ml or  
 CC 500 pg/ml, respectively, from Peripheral Blood Mononuclear Cells (PBMC)  
 CC withdrawn from TB patients, or PP positive individuals, 6 months after  
 CC diagnosis. (d) PP induces a specific antibody response in a TB patient,  
 CC as determined by enzyme linked immunosorbent assay (ELISA) technique or  
 CC a western blot. (e) PP induces a positive delayed type hypersensitivity  
 CC (DTH) response, determined by intradermal injection. (I) and (II) are  
 CC useful in preparing a prophylactic or therapeutic medicine as a vaccine  
 CC for induction of a protective or generation of an immune response in a  
 CC mammal against infection with a virulent Mycobacterium. (I) and (II) are  
 CC also useful as diagnostic reagent for the diagnosis of a virulent  
 CC Mycobacterium infection. The vaccine of the invention induces efficient  
 CC immunological memory, providing long term protection against TB. This

CC sequence represents a Microbacterium tuberculosis TB54 antigen  
 CC described in the invention.  
 XX  
 SQ Sequence 495 AA;  
 Alignment Scores:  
 Pred. No.: 1,66e-128 Length: 495  
 Score: 1550.50 Matches: 297  
 Percent Similarity: 76.52% Conservative: 68  
 Best Local Similarity: 62.26% Mismatches: 109  
 Query Match: 56.65% Indels: 3  
 DB: 21 Gaps: 3  
 US-09-759-990A-1 (1-1461) x AAY87873 (1-495)  
 QY 34 CAGTACAGAAATGCCGATCAATCAACTCCATGTTCTCGGCCGTAAGGAACCTTACCTTGCT 93  
 Db 20 AspPheLysIleAlaAspLeuSerLeuAlaAspPheGlyArgLysGluLeuArgIleAla 39  
 QY 94 CAGAAAGAAATGCCAGGCTCTTATGGTTCTTCTGTGACGTTATTCGGCTTCTTAAGCCATTG 153  
 Db 40 GluHisGluMetProGlyLeuMetSerLeuArgArgGluTyrAlaGluValGlnProLeu 59  
 QY 154 AAGGCTCTCAGAAATCTCTGTTCCCTCCATCAGATCCAGACAGCCGCTCTCATCCGAG 213  
 Db 60 LysGlyAlaArgIleSerGlySerLeuHisMetThrValGlnThrAlaValLeuIleGlu 79  
 QY 214 ACATCTCAGCTCTTGGTGTCTGATGTCTGATGGCTTCTGCAATCTCTTCTACACAA 273  
 Db 80 ThrLeuThrAlaLeuGlyAlaGluValArgTrpAlaSerCysAsnIlePheSerThrGln 99  
 QY 274 GATACAGCCGCTGCTGCTATGTTCTCGGCCGCAACAGGCGACACAGAGCCAGCCGCT 333  
 Db 100 AsphHisAlaAlaAlaValValGlyProHisGlyThrProAspGluProLysGly 119  
 QY 334 ATCCCACTCTTCCCTCGAAGGCGCAACACTCCAGATACTCGGGAACACATACCCG 393  
 Db 120 ValProValPheAlaTrpLysGlyGluThrLeuGluGluTyrTrpTrpAlaAlaGluGln 139  
 QY 394 GCTCTCATGTGCGCAGATGTCTCAAGGCCCA---CAGCAGTTGTCGATGTGTGTGAT 450  
 Db 140 MetLeuThrTrpProAspProAspLysProAlaAsnMetIleLeuAspAspGlyGlyAsp 159  
 QY 451 GCTACACTCTCTCATCTCAAGGCTTCGAATTCGAACAGCGGCTGTGTTCAGAGCCA 510  
 Db 160 AlaThrMetLeuValLeuArgGlyMetGlnTrpGluLysAlaGlyValValProAla 179  
 QY 511 ACAGAACTCAGAACCTCGAATACCGCTGCTTCTTCTTACACTCAAGCAGGCTTCTCAAC 570  
 Db 180 GluGluAspAspProAlaGluTrpLysValPheLeuAsnLeuLeuArgThrArgPheGlu 199  
 QY 571 CAAGACAGAACACCTGGGCACACAGTTGCTCCCGGATGCAACGGTGTTCGAGAGACA 630  
 Db 200 ThrAspLysAspLysTrpThrLysIleAlaGluSerValLysGlyValThrGluGluThr 219  
 QY 631 ACACAGAGTGTCCACCGCTCTTACAGCTCGAAGAGGCGCAAACTCTCTTCCAGCC 690  
 Db 220 ThrThrGlyValLeuArgLeuTyrGlnPheAlaAlaGlyAspLeuAlaPheProAla 239  
 QY 691 ATCAAGCTCAACGACCTGTTCAAAAGTTCGAAGTTCGATAACATCTACGGCTCGCCGAC 750  
 Db 240 IleAsnValAsnAspSerValThrLysSerLysPheAspAsnLysTyrGlyThrArgHis 259  
 QY 751 TCCCTTATCATGATCAACCGCTCTTCCGATGTCTATGTCGGCGGCAACAGCAGCTCTC 810  
 Db 260 SerLeuIleAspGlyIleAsnArgGlyThrAspAlaLeuIleGlyGlyLysValLeu 279  
 QY 811 GTCATGGGTTACCGGATCTCGCAAGGGCTGCGCTCAATCCCTCGTGGCCAGGCGCT 870  
 Db 280 IleCysGlyTyrGlyAspValGlyLysGlyCysAlaGluAlaMetLysGlyGlnGlyAla 299  
 QY 871 CGCGTTATCATCAGAACTCGACCCCAATCTGCGCTCTCCAGCTGCCATCGAAGGCTAC 930  
 Db 871 CGCGTTATCATCAGAACTCGACCCCAATCTGCGCTCTCCAGCTGCCATCGAAGGCTAC 930

```
Db      300 ArgValSerValThrGlulLeuLeuProLeuAsnAlaLeuGlnAlaValMetMetGluGlyPhe 319
Qy      931 CAGGTCGCCCGCATCGAGGAAGTCGTCGAAGGATCGATATCTTCGTACATGCACAGGA 990
Db      320 AspValValThrValGluGluAlaIleGlyAspAlaAspIleValValThrAlaThrGly 339
Qy      991 AACTCGCATATCTCTGTCGACATGATGCCCGCAGATGAAGATGAAGCTATTCGTGGT 1050
Db      340 AsnLysAspIleIleMetLeuGluHisIleLysAlaMetLysAspHisAlaIleLeuGly 359
Qy      1051 AACATCGCCACTTCGATACCAAAATGATACAGATGGCTCATGAATACCAAGCATC 1110
Db      360 AsnIleGlyHisPheAspAsnGluLeuSerMetAlaGlyLeuGluArg---SerGlyAla 378
Qy      1111 AGGCACATCCCAATCAGCAGCAATACGACATGTGGGAATCCAGAT---GGCCACGCT 1167
Db      379 ThrArgValAsnValLysProGlnValAspLeuThrPheGlyAspThrGlyArgSer 398
Qy      1168 ATCCCTCTCTGTCGAGGCGCGCTCTTAACCTTGGCTGCGCTACAGGTCAACCATCT 1227
Db      399 IleIleValLeuSerGluGlyArgLeuLeuAsnLeuGlyAsnAlaThrGlyHisProSer 418
Qy      1228 TTCGTATGTCAATGTCAATCAACAACAGACACTCGCTCAGCTCGACCTTACGAAAG 1287
Db      419 PheValMetSerAsnSerPheAlaAsnGlnThrIleAlaGlnIleGluLeuThrThrLys 438
Qy      1288 AGAGAAATCTCGAGAAGAGGTTTACACACTTCGGAAGCATCTCGAAGCATCTGATGAAGAAGTCGCT 1347
Db      439 AsnAspGluTyrAspAsnGluValTyrArgLeuProLysHisLeuAspLulLysValAla 458
Qy      1348 CGCCTCCACCTCGCATCTCTCGATGTCACACTTCAAGCTTACAAAGCTTACAGAGCGGCTGAC 1407
Db      459 ArgIleHisValGluAlaLeuGlyGlyHisLeuThrLysLeuThrLysGluGlnAlaGlu 478
Qy      1408 TACATCAAGTTCAGTGAAGGTCTTCATCAAGTCTGATGCTTACCGTTAT 1458
Db      479 TyrLeuGlyValAspValGluGlyProTyrLysProAspHisTyrArgTyr 495

RESULT 3
AAG40086
ID      AAG40086 standard; Protein; 485 AA.
AC      AAG40086;
XX
XX
DT      18-OCT-2000 (first entry)
XX
DE      Arabidopsis thaliana protein fragment SEQ ID NO: 49691.
KW      Protein identification; signal transduction pathway; metabolic pathway;
KW      hybridisation assay; generic mapping; gene expression control; promoter;
KW      termination sequence.
XX
OS      Arabidopsis thaliana.
XX
XX      BP1033405-A2.
XX
XX      06-SEP-2000.
XX
XX      25-FEB-2000; 2000EP-0301439.
XX
XX      25-FEB-1999; 99US-0121825.
XX      05-MAR-1999; 99US-0123160.
XX      09-MAR-1999; 99US-0123548.
XX      23-MAR-1999; 99US-0125788.
XX      29-MAR-1999; 99US-0126264.
XX      29-MAR-1999; 99US-0126785.
XX      01-APR-1999; 99US-0127462.
XX      06-APR-1999; 99US-0128234.
XX      16-APR-1999; 99US-0128714.
XX      16-APR-1999; 99US-0129845.
XX      19-APR-1999; 99US-0130077.
XX      21-APR-1999; 99US-0130449.
XX      23-APR-1999; 99US-0130510.
XX      23-APR-1999; 99US-0130891.
XX      28-APR-1999; 99US-0131449.
XX      30-APR-1999; 99US-0132048.
XX      30-APR-1999; 99US-0132407.
XX      04-MAY-1999; 99US-0132484.
XX      05-MAY-1999; 99US-0132485.
XX      06-MAY-1999; 99US-0132486.
XX      06-MAY-1999; 99US-0132487.
XX      07-MAY-1999; 99US-0132863.
XX      11-MAY-1999; 99US-0134256.
XX      14-MAY-1999; 99US-0134218.
XX      14-MAY-1999; 99US-0134219.
XX      14-MAY-1999; 99US-0134221.
XX      14-MAY-1999; 99US-0134370.
XX      18-MAY-1999; 99US-0134768.
XX      19-MAY-1999; 99US-0134941.
XX      20-MAY-1999; 99US-0135124.
XX      21-MAY-1999; 99US-0135353.
XX      24-MAY-1999; 99US-0135629.
XX      25-MAY-1999; 99US-0136021.
XX      27-MAY-1999; 99US-0136392.
XX      28-MAY-1999; 99US-0136782.
XX      01-JUN-1999; 99US-0137222.
XX      03-JUN-1999; 99US-0137528.
XX      04-JUN-1999; 99US-0137502.
XX      07-JUN-1999; 99US-0137724.
XX      08-JUN-1999; 99US-0138094.
XX      10-JUN-1999; 99US-0138540.
XX      10-JUN-1999; 99US-0138847.
XX      14-JUN-1999; 99US-0139119.
XX      16-JUN-1999; 99US-0139452.
XX      16-JUN-1999; 99US-0139453.
XX      17-JUN-1999; 99US-0139492.
XX      18-JUN-1999; 99US-0139454.
XX      18-JUN-1999; 99US-0139455.
XX      18-JUN-1999; 99US-0139456.
XX      18-JUN-1999; 99US-0139457.
XX      18-JUN-1999; 99US-0139458.
XX      18-JUN-1999; 99US-0139750.
XX      18-JUN-1999; 99US-0139763.
XX      21-JUN-1999; 99US-0139817.
XX      22-JUN-1999; 99US-0139899.
XX      23-JUN-1999; 99US-0140353.
XX      23-JUN-1999; 99US-0140354.
XX      24-JUN-1999; 99US-0140695.
XX      28-JUN-1999; 99US-0140823.
XX      29-JUN-1999; 99US-0140991.
XX      30-JUN-1999; 99US-0141287.
XX      01-JUL-1999; 99US-0141842.
XX      01-JUL-1999; 99US-0142154.
XX      02-JUL-1999; 99US-0142055.
XX      06-JUL-1999; 99US-0142390.
XX      08-JUL-1999; 99US-0142803.
XX      09-JUL-1999; 99US-0142920.
XX      12-JUL-1999; 99US-0142977.
XX      13-JUL-1999; 99US-0143542.
XX      14-JUL-1999; 99US-0143624.
XX      15-JUL-1999; 99US-0144005.
XX      16-JUL-1999; 99US-0144085.
XX      16-JUL-1999; 99US-0144086.
XX      19-JUL-1999; 99US-0144325.
XX      19-JUL-1999; 99US-0144331.
XX      19-JUL-1999; 99US-0144332.
XX      13-JUL-1999; 99US-0144333.
XX      19-JUL-1999; 99US-0144334.
XX      19-JUL-1999; 99US-0144335.
XX      20-JUL-1999; 99US-0144352.
XX      20-JUL-1999; 99US-0144632.
```

PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144844.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145122.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145991.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149388.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149829.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
PR 29-OCT-1999; 99US-0162142.

## Alignment Scores:

Pred. No.: 2 59e-127 Length: 485  
Score: 1537.00 Matches: 302  
Percent Similarity: 74.33% Conservative: 60  
Best Local Similarity: 62.01% Mismatches: 109  
Query Match: 56.16% Indels: 16  
DB: 21 Gaps: 4

US-09-759-990A-1 (1-1461) x AAG40086 (1-485)

QY 19 ACTGGTCTCCATTCCAGTACAGAAATCCGACATCAACCTCCATGTTCTCGGCCGTAAG 78  
Db 8 ThrSerGlyArgGluTyrLysValLysAspMetSerGlnAlaAspPheGlyArgLeu 27  
QY 79 GAATTTACCTTCTGCTCAGAGGAATCCAGGCTTATGTTCTTCGTCGAGCGTTATCC 138  
Db 28 GluLeuGluLeuAlaGluValGluMetProGlyLeuMetAlaCysArgThrGluPheGly 47  
QY 139 GCTTTAAGCCATTGAAGGCTGTGAGATCTCTGGTTCCCTCCATCATCAGTCCAGCA 198  
Db 48 ProSerGlnProPheLysGlyAlaArgIleThrGlySerLeuHisMetThrIleGlnThr 67  
QY 199 GCGTCTCTCATCCAGACTCATCAGCTCTTGGTCTGATGTGAGANGGCTTCTGCAAC 258  
Db 68 AlaValLeuIleGluThrLeuThrAlaLeuGlyAlaGluValArgTyrCysSerCysAsn 87  
QY 259 ATCTTCTCTACAGATACAGCGCTGCTGCTATCTGTTGCGGCCACAGGCACACCA 318  
Db 88 IlePheSerThrGlnAspHisAlaAlaAlaAlaAlaAlaArgSerAlaAla----- 105  
QY 319 GAGAAGCCAGCCGGTATCCAGTCTTCCCTGGAAGCGGGAACACTCCAGATACTGG 378  
Db 106 -----ValPheAlaTyrPlysGlyGluThrLeuGlnGluTyrTrp 118  
QY 379 GAGAACACATACCGGCTCTCACATGGCCAGATGTTCAAGCCGACAGAGTTGCGAT 438  
Db 119 TrpCysThrGluArgAlaLeuAspTrpGlyGlyGlyProAspLeuIleValAsp 138  
QY 439 GATGGTGTGATGCTACTCTCTCATCTCCAGAGGCTTCGAA-----TTCGAA 486  
Db 139 AspGlyGlyAspAlaThrLeuLeuIleHisGluGlyValLysAlaGluIlePheGlu 158  
QY 487 ACAGCCGGTCTGTTCCAGAGCCACAGAGCTGACAACTCGAATACCGTGGTTCCTT 546  
Db 159 LysThrGlyGlnValProAspProThrSerThrAspAsnProGluPheGlnIleValLeu 178  
QY 547 GCTACTACTCAGCAGGCTTTCACCAAGACAGAACCACTGCCACACAGTTGTCGCGC 606  
Db 179 SerIleIleLysGluGlyLeuGlnValAspProLysLysTyrHisLysMetLysGluArg 198  
QY 607 ATGAACGGTGTTCGAGAGAGACACACAGGTTCCACCGGCTCTACACCTCGAGAG 666



```
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 08-AUG-1999; 99US-0147260.
PR 08-AUG-1999; 99US-0147303.
PR 08-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
```

```
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

## Alignment Scores:

Pred. No.:	2,65e-127	Length:	508
Score:	1537.00	Matches:	302
Percent Similarity:	74.33%	Conservative:	60
Best Local Similarity:	62.01%	Mismatches:	109
Query Match:	56.16%	Indels:	16
DB:	21	Gaps:	4

US-09-759-990A-1 (1-1461) x AAG40085 (1-508)

```
Qy 19 ACTGGTCTCCATTCGAGTACAGAAATGCCGACATCAACCTCCATGTTCTCGGCCGTAAG 78
    :::::
Db 31 ThrSerSerGlyArgGluTyrLysValLysAspMetSerGlnAlaAspPheGlyArgLeu 50
    :::::
Qy 79 GAACCTTACCCTTCTCGAAGGAATGCCAGGTCCTATGTTCTTCGTGACGCGTTATTC 138
    :::::
Db 51 GluLeuGluLeuAlaGluValGluMetProGlyLeuMetAlaCysArgThrGluPheGly 70
    :::::
Qy 139 GCTTCTAAGCATTGAAGGTGTCAGAAATCTGTTCCCTCCATCCACATGATCCAGACA 198
    :::::
Db 71 ProSerGlnProPheLysGlyAlaArgIleThrGlySerLeuHisMetThrIleGlnIhr 90
    :::::
Qy 199 GCCCTCTCTCATCGAGACACTCACAGCTCTTGGTGCTGCTGATGCAGATGGCGCTTCCTGCAAC 258
    :::::
Db 91 AlaValLeuIleGluThrLeuThrAlaLeuGlyAlaGluValArgTyrCysSerCysAsn 110
    :::::
Qy 259 ATCTTCTCTACACAGATACAGCCGCTGCTGCTATGTTGTCGGCCCAACAGGCACACCA 318
    :::::
Db 111 IlePheSerThrGlnAspHisAlaAlaAlaIleAlaArgAspSerAlaAla----- 128
```



```
QY 319 GAGAGCCAGCCGGTATCCAGTCTTCGCGCTGGAAGGGGGAACACACTCCAGAACTACTGG 378
Db 129 -----ValPheAlaTrpLysGlyGluThrLeuGlnIuTrp 141
QY 379 GAGAACACATACCGGCTCTCATGCGCAGATGTCAGGCCACAGCAGCGTTCGAT 438
Db 142 TrpCysThrGluAgaAlaLeuAspTrpGlyProGlyGlyProAspLeuIleValAsp 161
QY 439 GATGCTGTGATGTCACATCTCAATCCAAAGGGCTTCGAA-----TTGAA 486
Db 162 AspGlyGlyAspAlaThrLeuLeuIleHisGluGlyValLysAlaGluGluIlePheGlu 181
QY 487 ACAGCGGCTGTCTTCAGAGCCACAGAACTGACAACTCGAATACCGCTGCGTCTT 546
Db 182 LysThrGlyGlnValProAspProThrSerThrAspAsnProGluPheGlnIleValLeu 201
QY 547 GGTACACTCAAGCGGCTCTTCAACCAAGACAAAGAACCACTGGCCACACAGCTTCTCCGGC 606
Db 202 SerIleIleLysGluGlyLeuGlnValAspProLysLysTyHisLysMetLysGluArg 221
QY 607 ATGAACGGTGTTCGGAAGAGACACACAGGTGTCCACCGCTCTACAGCTCGAGAAG 666
Db 222 LeuValGlyValSerGluGluThrThrThrGlyValLysArgLeuTyrglnMetGlnGln 241
QY 667 GAGGCAAAACCTCTTCCAGCCATCAACGTCACAGCCTGTTACAAAGTCCAGTTC 726
Db 242 AsnGlyThrLeuLeuPheProAlaIleAsnValAsnAspSerValThrLysSerLysPhe 261
QY 727 GATAACATCTACGGCTGCGCCACTCCCTTATCGATGATCAACCGCTCTTCGATGTC 786
Db 262 AspAsnLeuTyrglyCysArgHisSerLeuProAspGlyLeuMetArgAlaThrAspVal 281
QY 787 ATGATCGCGGCGACAGACAGCTCTGTCATGGGTACGGCATCTCGGCAAGGCTCGCCT 846
Db 282 MetIleAlaGlyLysValAlaValIleCysGlyTyrglyAspValGlyLysGlyCysAla 301
QY 847 CAATCCCTCCGCGCCAGCGCTGCGTTATCATCAAGAACTCGACCCCAATCTGGCT 906
Db 302 AlaAlaMetLysThrAlaGlyAlaArgValIleValThrGluIleAspProIleCysAla 321
QY 907 CTCAGGTCGCCATGGAAGCTACCAAGTCCGCGCATCGAGGAAGTCTCAAGATGTC 966
Db 322 LeuGlnAlaLeuMetGluGlyLeuGlnValLeuThrLeuGluAspValValSerGluAla 341
QY 967 GATATCTGTTATCATCAAGCAATCCCAATCAAGCAATCAAGCAATCAAGCAATGTTGG 1026
Db 342 AspIlePheValThrThrThrGlyAsnLysAspIleIleMetValAspHisMetArgLys 361
QY 1027 ATGAAGGATAAGGCTATTGTGCGTAACATCGCGCCACTTCGATACGAAATTCATACAGAT 1086
Db 362 MetLysAsnAsnAlaIleValCysAsnIleGlyHisPheAspAsnGluIleAspMetLeu 381
QY 1087 GGCCTCATGAATACCCAGGATCAAGCAATCCCAATCAAGCAATCAAGCAATGATGG 1146
Db 382 GlyLeuGluThrTyrglyValLysArgIleThrIleLysProGlnThrAspArgTrp 401
QY 1147 GAATTCACGAGTCCAGCT---ATCCCTCTTCTGCTGAGCGCGCTTCTTAACCTT 1203
Db 402 ValPheProGluThrLysAlaGlyIleIleValLeuAlaGlyArgLeuMetAsnLeu 421
QY 1204 GGCTCGCTACAGGTACCCATCTTCGTTATGTCAATGTCAATCAACCAACACAGACATC 1263
Db 422 GlyCysAlaThrGlyHisProSerPheValMetSerCysSerPheThrAsnGlnValIle 441
QY 1264 GCTCAGTCTGACCTCTACGAAAGAGA-----GGAATCTCGAAGAAGGTTTACACA 1317
Db 442 AlaGlnLeuGluLeuTrpAsnGluLysAlaSerGlyLysTyrglyLysLysValTyVal 461
QY 1318 CTTCCGAGCATCTCGATGAGAGTCCGCTCCCTCCACTCCGATCTCTCGATGTCAC 1377
Db 462 LeuProLysHisLeuAspGluLysValAlaLeuLeuHisLeuGlyLysLeuGlyAlaArg 481
QY 1378 CTTACAAAGCTTACACAGAAGCGGTGACTACATCAACGTTCCAGTTCAGGGTCCCTTAC 1437
```

```
Db 482 LeuThrLysLeuSerLysAspGlnSerAspTyrglnSerIleProIleGluGlyProTyrg 501
QY 1438 AAGTCTGATGCTTACCGTTAT 1458
Db 502 LysProProHisTyrgTyrgTyrg 508
RESULT 5
AAR26500
ID AAR26500 standard; Protein; 486 AA.
AC AAR26500;
XX
DT 10-MAR-1993 (first entry)
XX
DE Prod. of Nicotiana tabacum gene expressing at floral differentiation.
XX
KW Flower; induction.
XX
OS Nicotiana tabacum.
XX
PN JP04258292-A.
XX
PD 14-SEP-1992.
XX
PF 14-FEB-1991; 91JP-0020702.
XX
PR 14-FEB-1991; 91JP-0020702.
XX
PA (NISR) JAPAN TOBACCO INC.
XX
DR WPI; 1992-354683/43.
XX
DR N-PSDB; AAQ29419.
XX
PT Gene expressing at floral differentiation for flowering control -
PT obtd. from culture of floral axis epithelium cells of Nicotiana
PT tabacum, for introduction into other plants
XX
PS Disclosure; Page 6; 8pp; Japanese.
XX
CC The protein sequence was deduced from the DNA sequence of a clone
CC obtd. by screening a cultured epithelial axis cDNA library with probes
CC from cultured and untreated mRNA. The clone obtd. expressed its
CC protein only at floral differentiation. The gene can be introduced
CC into other plants or can be suppressed by an antisense technique for
CC the control of flowering of plants.
XX
SQ Sequence 486 AA;
Alignment Scores:
Pred. No.: 2,45e-126 Length: 486
Score: 1526.00 Matches: 304
Percent Similarity: 73.86% Conservative: 52
Best Local Similarity: 63.07% Mismatches: 110
Query Match: 55.75% Indels: 16
DE: 13 Gaps: 4
US-09-759-990A-1 (1-1461) x AAR26500 (1-486)
QY 34 GAGTACAGAAATTCGACATCAACCTCCATGTTCTCGCGCTAAGAACTTACCTTGTCT 93
Db 14 GluTyrlsValLysAspMetSerGlnAlaAspPheGlyArgLeuGluIleGluAla 33
QY 94 GAGAGGAATGCGAGTCTTATGTTCTTCGTCGAGCTTATTCGCTTCTTAAGCCATG 153
Db 34 GluValGluMetProGlyLeuMetAlaCysArgThrGluPheGlyProSerGlnProPhe 53
QY 154 AAGGTCGAGATCTCTGTTCTCCCTCCATCATCATCATCATCATCATCATCATCATCAT 213
Db 54 LysGlyAlaLysIleThrGlySerLeuHisMetThrIleGlnThrAlaValLeuIleGlu 73
QY 214 ACATCTCAGCTCTTGGTCTGATGTCAGATGGCTTCCTGCAACATCTTCTTACACAA 273
```

Db 74 ThrLeuThrAlaLeuGlyAlaGluValargTrpCysSerCysAsnIlePheSerThrGln 93  
Qy 274 GATACAGCGGTGCTGTATCGTGTGCGGCCAACAGACACACAGAGAACCGCGGT 333  
Db 94 AspHisAlaAlaAlaAlaAlaAlaArgAspSerAlaAla----- 106  
Qy 334 ATCCAGCTCTCCGCGTGGAGCGGCAACACTCCCAATACCTGGGAGAGACACATACGC 393  
Db 107 -----ValPheAlaTrpGlyGluThrLeuGlnGluTrpTrpCysThrGluArg 124  
Qy 394 GCTCTCATGCGCCAGATGCTCAAGGCCACAGACAGTGTCTCGATGATGGTGGTATGCT 453  
Db 125 AlaLeuAspTrpGlyProGlyGlyProAspLeuIleValAspGlyGlyAspAla 144  
Qy 454 ACATCTCTCTCTCCAGGCGCTC-----GAATCGAAACAGCGGTGCTGTT 501  
Db 145 ThrLeuLeuIleHisGluGlyVallysAlaGluGluPheAlaLysAsnGlyThrIle 164  
Qy 502 CCAGAGCCAAACAGAGCTGACACCTCGAATACCGTGGCTCTCTGCTACACTCAGCAG 561  
Db 165 ProAspProAsnSerThrAspAlaGluPheGlnLeuValLeuThrIleIleGln 184  
Qy 562 GTCCTTCAACCAAGACCAACACTGGCACACAGTGTCTCGCGCATGAACGGTGTTC 621  
Db 185 SerLeuLysThrAspProLeuLysTrpThrLysMetLysGluArgLeuValGlyValSer 204  
Qy 622 GAAGAGACAAACAGGTGTCCACCGCTCTACAGCTCGAGAGAGGGGCAACTCTC 681  
Db 205 GluGluThrThrThrGlyVallysArgLeuTrpGlnMetGlnAlaAsnGlyThrLeuLeu 224  
Qy 682 TTCCAGCGCATCAACGTCACAGAGCTGTTCAAGTCCAAAGTTCGATACACTATCAGGC 741  
Db 225 PheProAlaIleAsnValAsnAspSerValThrLysSerLysPheAspLeuMetGly 244  
Qy 742 TGCCGCCACTCCCTTATTCGATGTGATCAACCGCTCTCCGATGTCTATGATCGCGGCAAG 801  
Db 245 CysArgHisSerLeuProAspGlyLeuMetArgAlaThrAspValMetIleAlaGlyLys 264  
Qy 802 ACAGCTCTCTGTCAGGTACGGCATGTCCGCGAGGCTCGCTCAATCCCTCGGTGC 861  
Db 265 ValAlaLeuValAlaGlyTrpGlyAspValGlyLysGlyCysAlaAlaLeuLysGln 284  
Qy 862 CAAGCGCTCCGCTTATTCATCAGACAGACTCGACCAATCTCGCTCTCCAGGCTGCCATG 921  
Db 285 AlaGlyAlaArgValIleValThrGluIleAspProIleCysAlaLeuGlnAlaThrMet 304  
Qy 922 GAAGCTACCAAGTCCCGCATTCGAGGAGTCTCGAAGGATGTGATATCTCTGTACA 981  
Db 305 GluGlyLeuGlnValLeuThrLeuGluAspValValSerAspValAspIlePheValThr 324  
Qy 982 TGCACAGAACTCGGATATCTCTCTGACATGTGCGCCAGATGAGGATAGGCT 1041  
Db 325 ThrThrGlyAsnLysAspIleIleMetValAspHisMetArgLysMetLysAsnAla 344  
Qy 1042 ATTGTCGGTAACTCGGCCACTTCGATAACGAATTCATACAGATGGCTCATGAATAC 1101  
Db 345 IleValCysAsnIleGlyHisPheAspAsnGluIleAspMetLeuGlyLeuGluThrTyr 364  
Qy 1102 CCAGCATCAGACATCCCAATCAAGCCAGAAATACACATGTGGAAATCCAGATGGC 1161  
Db 365 ProGlyValLysLysIleThrIleLysProGlnThrAspArgTrpValPheProAspThr 384  
Qy 1162 CAGCGT---ATCTCTCTCTGCGGGCGGCTCTTAACCTTGGCTGGCTACAGGT 1218  
Db 385 AsnSerGlyIleIleValLeuAlaGluGlyArgLeuMetAsnLeuGlyCysAlaThrGly 404  
Qy 1219 CACCATCTTTCGTTATGTCAATGTCATTCACAAACAGACACTCGCTCAGCTCGACCTC 1278  
Db 405 HisProSerPheValMetSerCysSerPheThrAsnGlnValIleAlaGlnLeuGluLeu 424  
Qy 1279 TCAGAAAGAGA-----GGAATCTCGAGAGAGGTTTACACTCCGAGCACTC 1332  
Db 425 TrpAsnGluLysSerSerGlyLysTrpGluLysLysValTyrValLeuProHisLeu 444

Qy 1333 GATGAGAAGTGCCTCGCTCCACCTCGATCTCTCGATGCCACTTACAAAGCTTACA 1392  
Db 445 AspGluLysValAlaAlaAlaLeuHisLeuGlyLysLeuGlyValAlaLysLeuThrLysLeuSer 464  
Qy 1393 CAGAGAGAGCTGACTACACTACACTCAACCTTCAGTTCCAGGTCCTTACAGTCTGATGCTTAC 1452  
Db 465 LysAspGlnAlaAspTyrIleSerValProValGluGlyProTyrLysProAlaHisTyr 484  
Qy 1453 CTTTAT 1458  
Db 485 ArgTyr 486  
RESULT 6  
AAG34060  
ID AAG34060 standard; Protein; 485 AA.  
XX  
AC AAG34060;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Zea mays protein fragment SEQ ID NO: 41387.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence; corn.  
XX  
OS Zea mays subsp. mays.  
XX  
FN EP1033405-A2.  
XX  
ED 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 03-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140931.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142330.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-01433542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147392.  
PR 06-AUG-1999; 99US-0147360.  
PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 21-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 26-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:  
Pred. No.:  
Score:

4.52e-126  
1523.00

Length:  
Matches:

485  
301

Db	319	AspIlePheValThrThrGlyAsnLysAspIleIleMetValAspHisMetArgLys	336
QY	1027	ATGAGAGATAAGGCTATTGTCGGTAAATCATCGGCCCATCTCGATAACGAATAATGATACAGAT	1086
Db	339	MetLysAsnAsnAlaIleValCysAsnIleGlyHisPheAspAsnGluIleAspMetLeu	358
QY	1087	GGCCTCATGAAATACCCAGAGCATCAAGCACATCCCAATCAAGCCAGATACACATGTGG	1146
Db	359	GlyLeuGluThrTyProGlyValLysArgIleThrIleLysProGlnThrAspArgTyr	378
QY	1147	GAATTCCCAAGATGGCCAC--GGTATCTCTCTCTTGTGAGGGCCGCCCTCTTAACTTT	1203
Db	379	ValPheProGluThrAsnThrGlyIleIleValLeuAlaGluGlyArgLeuMetAsnLeu	398
QY	1204	GGCTGGCTACAGGTCACCCATCTTTCGTATTGTCAAATGTCATTCAAAACACAGACACTC	1263
Db	399	GlyCysAlaThrGlyHisProSerPheValMetSerCysSerPheThrAsnGlnValIle	418
QY	1264	GCTCACTCGACTCTACGAAAGAGA-----GGAATCTTCGAGAGAAGGTTTACACA	1317
Db	419	AlaGlnLeuGluLeuTyrLysGluLysSerSerGlyLysTyrGlyLysLysValTyrVal	438
QY	1318	CTTCGGAACATCTCGATGAAGAAGTCGCTCGCTCCACCTCGATCTCTCCATGTCAC	1377
Db	439	LeuProLysHisLeuAspGluLysValAlaAlaLeuHisLeuGlyLysLeuGlyAlaLys	458
QY	1378	CTTACAAACCTTACACAGACAGGCTGACTACATCAACGTTCCAGTTGAGGTCCTTAC	1437
Db	459	LeuThrLysLeuThrLysSerGlnAlaAspTyrIleSerValProIleGluGlyProTyr	478
QY	1438	AAGCTGTGATGTTACCGTTAT	1458
Db	479	LysProAlaHisTyrArgTyr	485
RESULT 7			
AAG34059			
ID	AAG34059 standard; Protein; 510 AA.		
XX	AAG34059;		
XX	18-OCT-2000 (first entry)		
XX	Zea mays protein fragment SEQ ID NO: 41386.		
DE	Zea mays protein fragment SEQ ID NO: 41386.		
XX	Protein identification; signal transduction pathway; metabolic pathway;		
XX	hybridisation assay; genetic mapping; gene expression control; promoter;		
XX	termination sequence; corn.		
CS	Zea mays subsp. mays.		
PN	EP1033405-A2.		
XX	06-SEP-2000.		
XX	25-FEB-2000; 2000EP-0301439.		
XX	25-FEB-1999; 99US-0121825.		
XX	05-MAR-1999; 99US-0123180.		
XX	09-MAR-1999; 99US-0123548.		
XX	23-MAR-1999; 99US-0125788.		
XX	25-MAR-1999; 99US-0126264.		
XX	29-MAR-1999; 99US-0126785.		
XX	01-APR-1999; 99US-0127462.		
XX	06-APR-1999; 99US-0128234.		
XX	08-APR-1999; 99US-0128714.		
XX	16-APR-1999; 99US-0128845.		
XX	19-APR-1999; 99US-0130077.		
XX	21-APR-1999; 99US-0130449.		
XX	23-APR-1999; 99US-0130510.		
XX	23-APR-1999; 99US-0130891.		
XX	28-APR-1999; 99US-0131449.		
XX	30-APR-1999; 99US-0132048.		
XX	30-APR-1999; 99US-0132407.		



```

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 4,63e-126
Score: 1523.00
Percent Similarity: 73.31%
Best Local Similarity: 61.81%
Query Match: 55.64%
Indels: 16
Gaps: 4
DB: 21

US-09-759-990a-1 (1-1461) x AAG34059 (1-510)

QY 19 ACTGGTGTCCATTGCGATCAGAAATGCGGACATCAACCTCCATGTTCTCGCGGTAAAG 78
DB 33 ThrSerSerGlyArgGluTyrLysValLysAspLeuSerGlnAlaAspPheGlyArgLeu 52
QY 79 GACTTACCTTCGTCTGAGAGGAATGCGAGCTCTTATGTTCTTCTGAGGCTTATTC 138
DB 53 GluIleGluLeuAlaGluValGluMetProGlyLeuMetAlaCysArgAlaGluPheGly 72
QY 139 GCTTCTAAGCCATTGAGGGTGCAGAACTCTGCTTCCCTCCACATGACAGTCCAGACA 198
DB 73 ProSerLysProPheAlaGlyAlaArgIleSerGlySerIleHisMetThrIleGlnThr 92
QY 199 GCGTCTCTCATGACACTCACAGCTCTTGTGTGATGTCAGATGGCTTCCTGCAAC 258
DB 93 AlaValLeuIleGluThrLeuThrAlaLeuGlyAlaGluValArgTyrCysSerCysAsn 112
QY 259 ATCTTCTCTACAGATACAGCGCTGCTGCTATGCTTGTGCTGCGCCACACAGCACCA 318
DB 113 IlePheSerThrGlnAspHisAlaAlaAlaIleAlaArgAspSerAlaAla----- 130
QY 319 GAGAACCCAGCGGTATCCAGTCTTCCGCTCGAAGGGCGAAACACTCCGAGAAATCTGG 378
DB 131 -----ValPheAlaTyrLysGlyGluThrLeuGluGluTyrTrp 143
QY 379 GAGAACCATACCGGCTCTCATGATGGCCAGATGTCAAGCCCCACAGAGTTCGAT 438
DB 144 TrpCysThrGluArgCysLeuAspTrpGlyGluAlaGlyGlyProAspLeuIleValAsp 163
QY 439 GATGGTGTGATGTACACTCCTCATCTCCAGGGCTTC-----GAATTCCAA 486
DB 164 AspGlyGlyAspAlaThrLeuLeuIleHisGluGlyValLysAlaGluGluTyrGlu 183
QY 487 ACAGCGGTGTGTTCCAGAGCCACAGAGCTGACACCTCGAATACGCTGCGTTCCTT 546
DB 184 LysThrGlyLysIleProAspProGluSerThrAspAsnAlaGluPheIleValLeu 203
QY 547 GCTACACTCAGAGGTCTTCAACACAGACAGACCACTGCGCACAGTGTGTCGCGG 606
DB 204 ThrIleIleArgAspGlyLeuLysAlaAspProLysLysTyrArgLysMetLysGluArg 223
QY 607 ATGACCGGTGTGTTCCGAGAGACACCAACAGAGTGTCCACCGCTCTTACAGCTCGAGAG 666
DB 224 LeuValGlyValSerGluThrThrThrGlyValLysArgLeuTyrGlnMetGlnGlu 243
QY 667 GAGGGCAACTGCTCTTCCAGGCATCAAGCTCAAGCAGCGTGTTCAAAGTCCAGTTC 726

```

```

DB 244 ThrGlyAlaLeuLeuPheProAlaIleAsnValAsnAspSerValThrLysSerLysPhe 263
QY 727 GATACACATCTACGGCTGCGGCACTCCCTTATCGATGCTATCAACCGTCTCCGATGTC 786
DB 264 AspAsnLeuTyrGlyCysArgHisSerLeuProAspGlyLeuMetArgAlaThrAspVal 283
QY 787 ATGATCGCGGCAAGACAGCTCTCTCATGGTTACGGCGATGTCGGCAAGGGCTCGCT 846
DB 284 MetIleAlaGlyLysValAlaValValCysGlyTyrGlyAspValGlyLysGlyCysAla 303
QY 847 CAATCCCTCGTGGCCAGGCGCTCGGCTTATCATCACAAGACTCGACCCCAATTCGGCT 906
DB 304 AlaAlaLeuLysGlnAlaGlyAlaArgValIleValThrGluIleAspProIleCysAla 323
QY 907 CTCAGGCTGCCATGGAAGGCTACAGGTCGCGCGCATCGAGGAAGTCGTCAGAGATGTC 966
DB 324 LeuGlnAlaLeuMetGluGlyLeuGlnValLeuProLeuGluAspValValSerGluAla 343
QY 967 GATATCTTGTATGTCACAGGAACCTCGATATCATCTCTGTGACATGATGGCCGAC 1026
DB 344 AspIlePheValThrThrThrGlyAsnLysAspIleIleMetValAspHisMetArgLys 363
QY 1027 ATGAGGATGAAGCTATTGTCGGTAAACATCGGCCACTTCGATAACGAAATTGATACAGT 1086
DB 364 MetLysAsnAsnAlaIleValCysAsnIleGlyHisPheAspAsnGluIleAspMetLeu 383
QY 1087 GGCCTCATGAATATACCAGGATCAAGCATCAACATCCCAATCAAGCCAGAAATACGATGG 1146
DB 384 GlyLeuGluThrTyrProGlyValLysArgIleThrIleLysProGlnThrAspArgTrp 403
QY 1147 GAATCCCGATGGCCAC---GCTATCTCTTCTGTGAGGGCGGCTTCTTACCTT 1203
DB 404 ValPheProGluThrAsnThrGlyIleIleValLeuAlaGluGlyArgLeuMetAsnLeu 423
QY 1204 GGCTCGCTACAGTCAACCATCTTCTGTTATGTCATATGTCATTCATTCACAAACCCAGACTC 1263
DB 424 GlyCysAlaThrGlyHisProSerPheValMetSerCysSerPheThrAsnGlnValIle 443
QY 1264 GCTCAGTCTGACCTTCACGAAAGAGA-----GGAATCTCGAGAGAGAGTTTACACA 1317
DB 444 AlaGlnLeuGluLeuTrpLysGluLysSerSerGlyLysTyrGluLysLysValTyrVal 463
QY 1318 CTTCCGAGGATCTCGATGAAGAGTGGCTCGCTCCACTCGGATCTCTCGATGTCAC 1377
DB 464 LeuProLysHisLeuAspGluLysValAlaAlaLeuHisLeuGlyLysLeuGlyAlaLys 483
QY 1378 CTTCAAAAGTTTACACAGAGCGCTGACTACATCAACGTTTCCAGTTGAGGGTCTTAC 1437
DB 484 LeuThrLysLeuThrLysSerGlnAlaAspTyrIleSerValProIleGluGlyProTyr 503
QY 1438 AAGTCTGATCTTACCGTTAT 1458
DB 504 LysProAlaHisTyrArgTyr 510

RESULT 8
AAG40087
ID AAG40087 standard; Protein; 467 AA.
XX
AC AAG40087;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 49692.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX

```

PD 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 29-MAR-1999; 99US-0126284.  
PR 01-APR-1999; 99US-0126785.  
PR 06-APR-1999; 99US-0127462.  
PR 08-APR-1999; 99US-0128234.  
PR 16-APR-1999; 99US-0128714.  
PR 19-APR-1999; 99US-0128845.  
PR 21-APR-1999; 99US-0130077.  
PR 23-APR-1999; 99US-0130449.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 24-MAY-1999; 99US-0135353.  
PR 25-MAY-1999; 99US-0136229.  
PR 27-MAY-1999; 99US-0136021.  
PR 28-MAY-1999; 99US-0136392.  
PR 01-JUN-1999; 99US-0136782.  
PR 03-JUN-1999; 99US-0137282.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137502.  
PR 08-JUN-1999; 99US-0137724.  
PR 10-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138687.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 28-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140891.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 26-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.

PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 08-OCT-1999; 99US-0157753.  
 PR 08-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 26-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

## Alignment Scores:

Pred. No.: 1e-125  
 Score: 1519.00  
 Percent Similarity: 75.11%  
 Best Local Similarity: 63.40%  
 Query Match: 55.50%  
 DB: 21

US-09-759-990A-1 (1-1461) x AAG40087 (1-467)

QY 70 GCGCTAAGGAACCTTCCCTGCTGAGAGGAATGCCAGGTCTTATGTTCTGCTGAG 129  
 Db 7 GlyArgLeuGluLeuGluGluValGluMetProGlyLeuMetAlaCysArgThr 26  
 QY 130 CATTATCCGCTTCAAGCAATGAGGGTGTCAAGATCTCTGGTTCCCTCCCATGACA 189  
 Db 27 GluPheGlyProSerGlnProPheGlyAlaArgIleThrGlySerLeuHisMetThr 46  
 QY 190 GTCCAGACAGCCCTTCATCAGACACTCAGAGCTTGTGCTGCTGATGAGGGCT 249  
 Db 47 IleGlnThrAlaValLeuIleGluThrLeuThrAlaLeuGlyAlaGluValArgTrpCys 66  
 QY 250 TCTGTCAACATCTTCTCTACCAAGATACAGCCGCTGCTGCTATGCTTCTGCGCCCAACA 309  
 Db 67 SerCysAsnIlePheSerThrGlnAspHisAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 86  
 QY 310 GGCACACAGAGAGCCAGCCGCTATCCAGCTTCCTGCGCTGGAAGGCGAACAACACTCCCA 369  
 Db 87 Ala-----ValPheAlaTrpLysGlyGluThrLeuGln 97  
 QY 370 GAATACTGGGAGACACATACCCGCTCTCAGATGGCCAGATGGTCAAGGCCACAGGAG 429  
 Db 98 GluTyrTrpTrpCysThrGluArgAlaLeuAspTrpGlyProGlyGlyGlyProAspLeu 117

QY 430 GTTCTCGATGATGGTGGTGTGATGCTACACTCTCTCATCTCCAGAGGCTTCGAA----- 480  
 Db 118 IleValAspAspGlyGlyAspAlaThrLeuLeuIleHisGlyValLeuAlaGluGlu 137  
 QY 481 ---TTCCAAACACGCGTCTGTTCAGAGCCACACAGAGCTGACACCTCGAATACCGC 537  
 Db 138 IlePheGluLysThrGlyGlnValProAspProThrSerThrAspAsnProGluPheGln 157  
 QY 538 TCGCTTCTTCTACTACTCAAGCAGGTCTTCAACCAAGACAGAACCACTGCACACAGTT 597  
 Db 158 IleValLeuSerIleIleLysGluGlyLeuGlnValAspProLysLysTyrHisLysMet 177  
 QY 598 GCTGCGGATGAACGGTGTTCGAAAGACACACACAGGTGTCCACCGCTCTACCAG 657  
 Db 178 LysGluArgLeuValGlyValSerGluGluThrThrGlyValLysArgLeuTyrGln 197  
 QY 658 CTCGGAAGGAGGCAACTCTCTCCAGCATCAAGCTCAACGCTCAACGCTGTGTACAAAG 717  
 Db 198 MetGlnAsnGlyThrLeuLeuPheProAlaIleAsnValAsnAspSerValThrLys 217  
 QY 718 TCCAAGTTCGATAACATCTACGGCTGCGCCACTCCCTTATCGATGGTATCAACCGTGT 777  
 Db 218 SerLysPheAspAsnLeuTyrGlyCysArgHisSerLeuProAspGlyLeuMetArgAla 237  
 QY 778 TCCGATCTCATGATCGCGGCGAAGACAGCTCTCTGTCATGGTTACGGCGATGTCGGAAG 837  
 Db 238 ThrAspValMetIleAlaGlyValAlaValIleCysGlyTyrGlyAspValGlyLys 257  
 QY 838 GGCTGCGCTCAATCCCTCGCGCCAGGCGCTCGGTTATCATCACAAGAACTCGACCCA 897  
 Db 258 GlyCysAlaAlaAlaMetLysThrAlaGlyAlaArgValIleValThrGluIleAspPro 277  
 QY 898 ATCTGCGCTCTCCAGCTGCCATGGAAGGTACAGGTCCGCGCATCGAGGAAGTGTCTC 957  
 Db 278 IleCysAlaLeuGlnAlaMetGluGlyLeuGlnValLeuThrLeuGluAspValVal 297  
 QY 958 AAGGATCTCGATATCTTGTACATGACACAGGAACCTGCGATCATCTCTGTGTGACATG 1017  
 Db 298 SerGluAlaAspIlePheValThrThrGlyAsnLysAspIleIleMetValAspHis 317  
 QY 1018 ATGGCCAGATGAAGGCTATTGTCTGTAACATCGGCCACTTCGATAACGAAAT 1077  
 Db 318 MetArgLysMetLysAsnAsnAlaIleValCysAsnIleGlyHisPheAspAsnGluIle 337  
 QY 1078 GATACAGATGCGCTCATGAATACCCAGGATCAACACATCCCAATCAACAGCCAGATAC 1137  
 Db 338 AspMetLeuGlyLeuGluThrThrProGlyValLysArgIleThrIleLysProGlnThr 357  
 QY 1138 GACATGTGGGAATTCACAGATGGCCAGCT---ATCTCTCTTCTGTGAGGGCCGCTT 1194  
 Db 358 AspArgTrpValPheProGluThrLysAlaGlyIleIleValLeuAlaGluGlyArgLeu 377  
 QY 1195 CTTAACTTGGCTGGCTACAGGTCAACCCATCTTTCGTTATGTCAATGTCAATTCACAAAC 1254  
 Db 378 MetAsnLeuGlyCysAlaThrGlyHisProSerPheValMetSerCysSerPheThrAsn 397  
 QY 1255 CAGACACTCCCTCAGCTCGACCTTACGAAAAGAGA-----GGAAATCTCGAGAGAG 1308  
 Db 398 GlnValIleAlaGlnLeuGluThrTrpAsnGluLysAlaSerGlyLysTyrGluLysLys 417  
 QY 1309 GTTACACACTCCGAGCATCTCGATGAAGAGTCCGCTCGCTCCACTCGGATCTCTC 1368  
 Db 418 ValTyrValLeuProLysHisLeuAspGluLysValAlaLeuLeuHisLeuGlyLysLeu 437  
 QY 1369 GATCTCCACTTCAAAAGCTTACACAGACAGGCTGACTCATCAACAGTTCAGTGTGAG 1428  
 Db 438 GlyAlaArgLeuThrLysLeuSerLysAspGlnSerAspTyrValSerIleProIleGlu 457  
 QY 1429 GCTCTTACAGTCTGATGCTTACCGTAT 1458  
 Db 458 GlyProTyrLysProHisTyrArgTyr 467

RESULT 9



AAG41407  
ID AAG41407 standard; Protein; 485 AA.  
XX AC AAG41407;  
XX DT 18-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 51514.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence.  
OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137794.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139422.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.

PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150894.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

## Alignment Scores:

Pred. No.: 4, 27e-125  
 Score: 1512.00  
 Percent Similarity: 73.72%  
 Best Local Similarity: 61.40%  
 Query Match: 55.24%  
 DB: 21  
 Gaps: 4

US-09-759-990a-1 (1-1461) x AAG41407 (1-485)

QY 19 ACTGGTCTTCATGACAGAAATGCCGACATCAACCTCCATGTTCTCGGCGTAAG 78  
 DB 8 ThrSerSerGlyArgGluTyrLysValLysAspMetSerGlnAlaAspPheGlyArgLeu 27  
 QY 79 GAACCTACCTTCCTGACAGAGAAATGCCAGGCTTATGCTTCTGAGCGTTATCC 138  
 DB 28 GluileGluLeuAlaGluValGluMetProGlyLeuValSerCysValThrGluPheGly 47

QY 139 GCTTCTAAGCCATTGAAGGTGTGAGAAATCTCTGCTTCCCTCCATGACAGTCCAGACA 198  
 DB 48 ProSerGlnProLeuLysGlyAlaArgIleThrGlySerLeuHisMetThrIleGlnThr 67  
 QY 199 GCCCTCTCATCGAGACTCACAGCTCTTGGTGTGATGTCAGATGGCTTCTCTCAAC 258  
 DB 68 AlaValLeuIleGluThrLeuThrAlaLeuGlyAlaGluValArgTyrCysSerCysAsn 87  
 QY 259 ATCTTCTCTACACAGATACAGCGCTGCTGCTATCGTTGTGCGGCCAACAGGCACACCA 318  
 DB 88 IlePheSerThrGlnAspHisAlaAlaAlaAlaAlaAlaArgAspSerAlaAla----- 105  
 QY 319 GAGAACCCAGCGGTATCCCGTTCCTGCTGGAAGGGGAAACACTCCCGAGATCTGG 378  
 DB 106 -----ValPheAlaTyrPlysGlyGluThrLeuGlnGluTyrTyr 118  
 QY 379 GAGAACACATACCGCGCTCTCACATGCCAGATGTCGACAGGCCACAGCAGGTTGTCAT 438  
 DB 119 TrpCysThrGluArgAlaLeuAspTyrGlyProGlyGlyGlyProAspLeuIleValAsp 138  
 QY 439 GATGGTGTGANGCTACTCTCTCATCTCCAGGGGCTTCGAA-----TTCGAA 486  
 DB 139 AspGlyGlyAspAlaThrLeuLeuIleHisGluGlyValLysAlaGluGluIlePheAla 158  
 QY 487 ACAGCGGTGCTGTCAGAGGCCAACAGAGCTGACACCTCGAATACCGCTGGGTCTT 546  
 DB 159 LysAsnGlyThrPheProAspProThrSerThrAspAsnProGluPheGlnIleValLeu 178  
 QY 547 GCTACTACTCAAGCAGGTCTTCAACCAAGACAGAACCACTGCGCACACAGTTGCTGCGGC 606  
 DB 179 SerIleIleLysAspGlyLeuGlnValAspProLysLysTyrHisLysMetLysGluArg 198  
 QY 607 ATGAACCGTGTTCGAGAGACAAACACAGGTGTCCACCGCTCTACAGCTCGAAG 666  
 DB 199 LeuValGlyValSerGluGluThrThrGlyValLysArgLeuTyrGlnMetGlnGlu 218  
 QY 667 GAGGCGAACTCTCTCCAGCCATCAAGCTCAAGCGCTGTACAAAGTCCAGATTC 726  
 DB 219 ThrGlyAlaLeuLeuPheProAlaIleAsnValAsnAspSerValThrLysSerLysPhe 238  
 QY 727 GATAACATCTACGGTTCGCGCACTCCCTTATGATGATGATCAACCGTGTTCGATGTC 786  
 DB 239 AspAsnLeuTyrGlyCysArgHisSerLeuProAspGlyLeuMetArgAlaThrAspVal 258  
 QY 787 ATGATCGCGCGAAGCAGCTCTCTCATGGTTCAGCGATGTCGGAAGGGCTGGCT 846  
 DB 259 MetIleAlaGlyLysValAlaValIleCysGlyTyrGlyAspValGlyLysGlyCysAla 278  
 QY 847 CAATCCCTCGTGGCCAGCGGCTCGGTTATCATCACAGAACTCGACCAATCTGGCT 906  
 DB 279 AlaIleMetLysThrAlaGlyAlaArgValIleValThrGluIleAspProIleCysAla 298  
 QY 907 CTCAGGCTGCCATGAAGGTTCACAGGTTCGCGCATCGAGGAGTGTCTCAAGGATGTC 966  
 DB 299 LeuGlnAlaLeuMetGluGlyLeuGlnValLeuThrLeuGluAspValValSerGluAla 318  
 QY 967 GATATCTCTTACATGACAGGAACTGCGATATCATCTCTGTTGACATGATGCGCCAG 1026  
 DB 319 AspilePheCysThrThrThrGlyAsnLysAspileMetValAspHisMetArgLys 338  
 QY 1027 ATGAGGATAAGCTATTGTTGTAACATCGCCACTTCGATAACGAAATGTATACAGAT 1086  
 DB 339 MetLysAsnAsnAlaIleValCysAsnIleGlyHisPheAspAsnGluIleAspMetLeu 358  
 QY 1087 GGCCTCATGAATAACCCAGGCATCAAGACATCCCAATCAAGCCAGAAATACGACATGG 1146  
 DB 359 GlyLeuGluThrTyrProGlyValLysArgIleThrIleLysProGlnThrAspArgTyr 378  
 QY 1147 GAATTCACCATGCGCACGCT---ATCCTCTTCTTCTGAGGCGCGCTTCTTACCTT 1203  
 DB 379 ValPheProAspThrAsnSerGlyIleValLeuAlaGluGlyArgLeuMetAsnLeu 398

```
QY 1204 GCGTCGGCTACAGGTCTCCCATCTTTCGTTATGCAATGTCATTCACAAACGACACTC 1263
D 1205 |||||
Db 399 GlyCysAlaThrGlyHisProSerPheValMetSerCysSerPheThrAsnGlnValIle 418
QY 1264 GCTCAGCTCGACCTCTACGAAAGAGA-----GGAATCTCGAGAGAGGTTTACACA 1317
D 1265 |||||
Db 419 AlaGlnLeuGluLeuTyrPheGlnLysSerSerGlyLysTyrGluLysValIleVal 438
QY 1318 CTTCCGAAGCATCTCGATGAAGAGTCGCTCGCTCCACCTCGCATCTCTCGATGCCAC 1377
D 1319 |||||
Db 439 LeuProLysHisLeuAspGluLysValAlaLeuHisLeuGlyLysLeuGlyAlaArg 458
QY 1378 CTTACAGACTTACAGAGAGCGCTGACTTACATCAAGTTCAGTTGAGGTCCTTAC 1437
D 1379 |||||
Db 459 LeuThrLysLeuThrLysAspGlnSerAspTyrValSerIleProValGluGlyProTyr 478
QY 1438 AAGCTCTGATCTTACCGCTTAT 1458
D 1439 |||||
Db 479 LysProValHisTyrArgTyr 485

RESULT 10
AAG41406
ID AAG41406 standard; Protein; 497 AA.
XX AC AAG41406;
XX
DI 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51513.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
```

PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 06-AUG-1999; 99US-0147360.  
PR 07-AUG-1999; 99US-0147403.  
PR 08-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 12-AUG-1999; 99US-0148319.  
PR 13-AUG-1999; 99US-0148341.  
PR 14-AUG-1999; 99US-0148565.  
PR 15-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 19-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 21-AUG-1999; 99US-0149929.  
PR 22-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 24-AUG-1999; 99US-0150566.  
PR 25-AUG-1999; 99US-0150584.  
PR 26-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 28-AUG-1999; 99US-0151080.  
PR 29-AUG-1999; 99US-0151303.  
PR 30-AUG-1999; 99US-0151438.  
PR 31-AUG-1999; 99US-0151930.  
PR 01-SEP-1999; 99US-0152363.  
PR 02-SEP-1999; 99US-0153070.  
PR 03-SEP-1999; 99US-0153758.  
PR 04-SEP-1999; 99US-0154018.  
PR 05-SEP-1999; 99US-0154039.  
PR 06-SEP-1999; 99US-0154779.  
PR 07-SEP-1999; 99US-0155139.  
PR 08-SEP-1999; 99US-0155486.  
PR 09-SEP-1999; 99US-0155659.  
PR 10-SEP-1999; 99US-0156458.  
PR 11-SEP-1999; 99US-0156596.  
PR 12-SEP-1999; 99US-0157117.  
PR 13-SEP-1999; 99US-0157153.  
PR 14-SEP-1999; 99US-0157885.  
PR 15-SEP-1999; 99US-0158029.  
PR 16-SEP-1999; 99US-0158232.  
PR 17-SEP-1999; 99US-0158369.  
PR 18-SEP-1999; 99US-0159293.  
PR 19-SEP-1999; 99US-0159294.  
PR 20-SEP-1999; 99US-0159295.  
PR 21-SEP-1999; 99US-0159329.  
PR 22-SEP-1999; 99US-0159330.  
PR 23-SEP-1999; 99US-0159331.  
PR 24-SEP-1999; 99US-0159637.  
PR 25-SEP-1999; 99US-0159638.  
PR 26-SEP-1999; 99US-0159584.  
PR 27-SEP-1999; 99US-0160741.  
PR 28-SEP-1999; 99US-0160767.  
PR 29-SEP-1999; 99US-0160768.  
PR 30-SEP-1999; 99US-0160770.  
PR 01-OCT-1999; 99US-0160814.  
PR 02-OCT-1999; 99US-0160815.  
PR 03-OCT-1999; 99US-0160980.  
PR 04-OCT-1999; 99US-0160981.  
PR 05-OCT-1999; 99US-0161404.  
PR 06-OCT-1999; 99US-0161405.  
PR 07-OCT-1999; 99US-0161406.  
PR 08-OCT-1999; 99US-0161359.  
PR 09-OCT-1999; 99US-0161360.  
PR 10-OCT-1999; 99US-0161361.  
PR 11-OCT-1999; 99US-0161920.

PR 28-OCT-1999; 99US-0161992.  
PR 29-OCT-1999; 99US-0161993.  
PR 30-OCT-1999; 99US-0162142.  
Alignment Scores:  
Pred. No.: 4.32e-125 Length: 497  
Score: 1512.00 Matches: 299  
Percent Similarity: 73.72% Conservative: 60  
Best Local Similarity: 61.40% Mismatches: 112  
Query Match: 55.24% Indels: 16  
DB: 21 Gaps: 4  
US-09-759-990A-1 (1-1461) x AAG41406 (1-497)  
Qy 19 ACTGGTCTCCATTTCAGTACAGAAATTCGCGACATCAACCTCCATGTTCTCGGCGGTAAAG 78  
Db 20 ThrSerSerGlyArgGluThrLysValLysAspMetSerGlnAlaAspPheGlyArgLeu 39  
Qy 79 GAACCTTACCTTCGCTGAGAGGAATGCCAGGTCTTATGGTTCTTCGAGCGGTATTC 138  
Db 40 GluIleGluLeuAlaGluValGluMetProGlyLeuValSerCysValThrGluPheGly 59  
Qy 139 GCTTCTAAGCATTGAAGGTGTGAGAAATCTCTGGTTCCTCCATCAGATGAGTCCAGAC 198  
Db 60 ProSerGlnProLeuLysGlyAlaArgIleThrGlySerLeuHisMetThrIleGlnThr 79  
Qy 199 GCGTCTCTCATCGAGACTCACAGCTTCTTGGTGTCTGATGATGAGTGGCTTCTCGAAC 258  
Db 80 AlaValLeuIleGluThrLeuThrAlaLeuGlyAlaGluValArgTrpCysSerCysAsn 99  
Qy 259 ATCTTCTCTACACAGATACAGCGCTGCTGCTATGTTGTCGCCCCACAGGCACACCA 318  
Db 100 IlePheSerThrGlnAspHisAlaAlaAlaIleAlaArgAspSerAlaAla----- 117  
Qy 319 GAGAAGCCAGCGGTATCCAGTCTTCGCTGGAAGGGGAAACACTCCAGATACTGG 378  
Db 118 -----ValPheAlaTrpLysGlyGluThrLeuGlnGluTrp 130  
Qy 379 GAGAACAATACCGCTCTCATATGCGCAGATGGTCAAGGCCACACAGGTGTGCTAT 438  
Db 131 TrpCysThrGluArgAlaLeuAspTrpGlyProGlyGlyProAspLeuIleValAsp 150  
Qy 439 GATGGTGTATCTACTCTCTCATCTCCAGAGGTCTCGAA-----TTCGAA 486  
Db 151 AspGlyGlyAspAlaThrLeuLeuIleHisGluGlyValLysAlaGluGluIlePheAla 170  
Qy 487 ACAGCCGGTCTTTCAGAGCCACAGAACTGCAACCTCGAATACCGTGGTCTT 546  
Db 171 LysAsnGlyThrPheProAspProThrSerThrAspAsnProGluPheGlnIleValLeu 190  
Qy 547 GCTACACTCAAGCAGGTCTTCAACCAAGACACCACTGGCAGACAGTGTGTCGGCG 606  
Db 191 SerIleIleLysAspGlyLeuGlnValAspProLysIleThrHisLysMetLysGluArg 210  
Qy 607 ATGAACGGTGTTCGAGAGACACACAGGTGTCCACCGCTCTACAGCTCGAAG 666  
Db 211 LeuValGlyValSerGluGluThrThrGlyValLysArgLeuThrGlnMetGlnGlu 230  
Qy 667 GAGGCAAACTCTCTTCCAGCCATCAAGCTCAAGCGTGTGTACAAAGTCAAGTTC 726  
Db 231 ThrGlyAlaLeuLeuPheProAlaIleAsnValAspSerValThrLysSerLysPhe 250  
Qy 727 GATACATCTACGGTCCGCGCTCTCCCTTATGATGGTATCAACCGTGTTCGGATTC 786  
Db 251 AspAsnLeuThrGlyCysArgHisSerLeuProAspGlyLeuMetArgAlaThrAspVal 270  
Qy 787 ATGATCGCGCGCAAGACAGCTCTCTGATGGTATCGCGGTATCGGCAAGGGTGGCT 846  
Db 271 MetIleAlaGlyLysValAlaValIleCysGlyThrGlyAspValGlyLysGlyCysAla 290  
Qy 847 CAATCCCTCCGTGGCCAGCGCTCGGTATCATCATCAAGACTCGACCACTCGCTCGCT 906  
Db 291 AlaAlaMetLysThrAlaGlyAlaArgValIleValThrGluIleAspProIleCysAla 310

QY 907 CTCAGGCTCCATGGAAGCTACAGGTCGCGCGCATGAGAGATGCTCAAGAGTGC 966  
 DB 311 LeuGlnAlaLeuMetGluGlyLeuGlnValLeuThrLeuGluAspValValSerGluAla 330  
 QY 967 GATATCTTCGTTACATGACAGAACTGCGATATCATCTCTGTTGACATATGCGCCAG 1026  
 DB 331 AspilePheCysThrThrGlyAsnLysAspilelelMetValAspHisMetArgLys 350  
 QY 1027 ATGAAGGATAAGGCTATTTCGGTAACATCGGCCACTTCGATAACGAAATGATACAGAT 1086  
 DB 351 MetLysAsnAsnAlaileValLysAsnileGlyHisPheAspAsnGluileAspMetLeu 370  
 QY 1087 GGCCTCATGAAATACCCAGGATCATCAGACATCCCAATCAAGCCGAGATACGATGTGG 1146  
 DB 371 GlyLeuGluThrTyProGlyValLysArgileThrileLysProGlnThrAspArgTyr 390  
 QY 1147 GAAATCCAGATGCCACGCT--ATCCTCCTTCTGCTGAGGCGGCTTCTTAACCTT 1203  
 DB 391 ValPheProAspThrAsnSerGlyileleValleuAlaGluGlyArgLeuMetAsnLeu 410  
 QY 1204 GCGTCGCTACAGGTCAACCCATCTTTCGTTATGTCAATGTCAATTCACAAACGACACTC 1263  
 DB 411 GlyCysAlaThrGlyHisProSerPheValMetSerCysSerPheThrAsnGlnValile 430  
 QY 1264 CTTCAAGCTCAGCTTACGAAAGAGA-----GGAAATCTCGAAGAGAGGTTTACACA 1317  
 DB 431 AlaGlnLeuGluLeuThrAsnGluLysSerSerGlyLysTyThrGluLysLysValTyVal 450  
 QY 1318 CTTCCGAGCATCTCGATGAGAAGTCGCTCGCTCCACTCGGATCTCTCGATGTCCAC 1377  
 DB 451 LeuProLysHisLeuAspGluLysValAlaAlaLeuHisLeuGlyLysLeuGlyAlaArg 470  
 QY 1378 CTTCAAGCTTACAGAGCAGGCTGACTACATCAACGTTCCAGTTTCAGGGTCCWTAC 1437  
 DB 471 LeuThrLysLeuThrLysAspGlnSerAspTyValSerIleProValGluGlyProTy 490  
 QY 1438 AAGTCTCATGCTTACCGTTAT 1458  
 DB 491 LysProValHisTyArgTyr 497  
 RESULT 11  
 AAW01459  
 ID AAW01459 standard; Protein; 485 AA.  
 XX  
 AC AAW01459;  
 XX  
 DT 23-FEB-1997 (first entry)  
 XX  
 DE Arabidopsis S-adenosyl-L-homocysteine hydrolase.  
 XX  
 KW S-adenosyl-L-homocysteine hydrolase; SHH; promoter; monooct; dicot;  
 KW transgenic plant; disease resistance; pathogen resistance.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN WO9632488-A1.  
 XX  
 PD 17-OCT-1996.  
 XX  
 PF 10-APR-1996; 96WO-GB00882.  
 XX  
 PF 10-APR-1995; 95GB-0007381.  
 XX  
 PR (ZENE ) ZENECA LTD.  
 XX  
 PA Draper J, Greenland AJ, Skipsey M, Warner S;  
 XX  
 PI WPI; 1996-477138/47.  
 XX  
 DR N-PSDB; AAT44515.  
 XX  
 PT S-adenosyl-L-homocysteine hydrolase promoter - used for driving  
 expression of effector genes, such as pathogen resistance genes, in

## PT transgenic plants

Example 4; Fig 6; 57pp; English.

CC Arabidopsis S-adenosyl-L-homocysteine hydrolase (SHH) (AAW01459),  
 CC catalyses the reversible hydrolysis of S-adenosyl-L-homocysteine  
 CC to adenosine and homocysteine. Its amino acid sequence was  
 CC deduced from the SHH gene (AAT44515) isolated from an Arabidopsis  
 CC genomic DNA library. The promoter of the Arabidopsis SHH gene is  
 CC useful for the expression of effector genes in transgenic plants.  
 XX

SQ Sequence 485 AA;

## Alignment Scores:

Pred. No.: 1,12e-123 Length: 485  
 Score: 1496.00 Matches: 295  
 Percent Similarity: 72.69% Conservative: 59  
 Best Local Similarity: 60.57% Mismatches: 117  
 Query Match: 54.66% Indels: 16  
 DB: 17 Gaps: 4

US-09-759-990A-1 (1-1461) x AAW01459 (1-485)

QY 19 ACTGGTCTCCATTCGAGTACAGAAATGCCGACATCAACCTCCATGTTCTCGCGCTAAG 78  
 DB 8 ThrSerSerGlyArgGlnTyLysValLysAspMetSerGlnAlaAspPheGlyArgLeu 27  
 QY 79 GAATTCATCCCTTCTGAGAGAAATGCCAGGTCTTATGTTCTTCGTGAGCGTTATTC 138  
 DB 28 GluLeuGluAlaGluValGluMetProGlyLeuMetAlaCysArgThrGluPheGly 47  
 QY 139 GCTTCTAAGCCATTGAAGGTGTGAGAAATCTCTGTTCTCCATCCATCAGTCCGACAC 198  
 DB 48 ProSerGlnAlaPheLysGlyAlaArgIleThrGlySerLeuHisMetThrIleGlnThr 67  
 QY 199 GCGTCTCTCATCGAGACACTCACAGCTCTTGTGCTGATGTCAGATGGCTTCCTGCAAC 358  
 DB 68 AlaValleuileGluThrLeuThrAlaLeuGlyAlaGluValArgTyrCysSerCysAsn 87  
 QY 259 ATCTTCTCTACACAGATACAGCGCTGCTGCTATCTGTTGTCGCCCAACAGGCACACCA 318  
 DB 88 IlePheSerThrGlnAspHisAlaAlaAlaAlaAlaAlaArgSerAlaAla----- 105  
 QY 319 GAGAACCCAGCCGGTATCCAGTCTTCGCTGGAAGGCGGAAACACTCCAGAAATCTGG 378  
 DB 106 -----AlaPheAlaTrpLysGlyGluThrLeuGlnGluTyTrp 118  
 QY 379 GAGAACACATACCGGCTCTCACATGCGCAGATGTCGAAGGCCACACAGCGTTGTCGAT 438  
 DB 119 TrpCysThrGluArgAlaLeuAspTrpGlyProGlyGlyProAspLeuileValAsp 138  
 QY 439 GATGGTGGTGTATACACTCTCTCATCTCCAAGGGTTCGAA-----TTCGAA 486  
 DB 139 AspGlyGlyAspAlaThrLeuLeuIleHisGluGlyValLysAlaGluGluIlePheGlu 158  
 QY 487 ACACCGGCTGCTTTCAGAGCCCAACAGAGCTGACAACTCGAATACCGTGGCTTCTT 546  
 DB 159 LysThrGlyGlnValProAspProThrSerThrAspAsnProGluPheGlnIleValLeu 178  
 QY 547 GCTACACTCAAGCAGCTCTTCAACCAAGACAAAGACCACTGGCACACAGTTGTCGCGGC 606  
 DB 179 SerIleleLysGluGlyLeuGlnValAspProLysTyThrLysMetLysGluArg 198  
 QY 607 ATGAACGGTGTTCGAGAGACACAAACAGGTGTCACCGCTCTACAGCTCGAGAAG 666  
 DB 199 LeuValGlyValSerGluGluThrThrGlyValLysArgLeuTyGlnMetGlnGln 218  
 QY 667 GAGGGCAAACTCTCTTCCAGCCCATCAACGTCAACAGCGCTGTTTACAACTCAAGTTC 726  
 DB 219 AsnGlyThrLeuLeuPheProAlaIleAsnValAsnAspSerValThrLysSerLysPhe 238  
 QY 727 GATAACACTCAGGCTGCGCCACTCCCTTATCGATGGTATCAACCGCTCTCCGATGC 786

```
Db      239 AspAsnLeuTyrGlySerGlyHisSerLeuProAspGlyLeuMetArgAlaThrAspVal 258
QY      787 ATGATCGCGCAGACAGACAGCTCTCGTCATGGTTACGGCGATCGGCAAGGGCTGCGCT 846
Db      259 MetIleAlaGlyLysValAlaValIleCysGlyTyrGlyAspValGlyLysGlyCysAla 278
QY      847 CAATCCCTCCGTCGCCAAGCGCTCGCTTATCATCAAGAACTCGACCAATCGCGCT 906
Db      279 AlaAlaMetLysThrAlaGlyAlaArgValIleValThrGluIleAspProIleCysAla 298
QY      907 CTCAGGCTGCCATGGAGGCTACAGGTCGCCGCGCATCGAGAGTCTGCAAGATGTC 966
Db      299 LeuGlnAlaLeuMetGluGlyLeuGlnValLeuThrLeuGluAspValValSerGluAla 318
QY      967 GATATCTTCGTATACACAGAACTCGCATATCATCTCTGTGACATATGCGCCAG 1026
Db      319 AspIlePheValThrThrGlyAsnLysAspIleIleMetValAspHisMetArgLys 338
QY      1027 ATGAGGATAGGCTATTGTCGTACATCGCCACTTCGATAAGCAATTCATACAGAT 1086
Db      339 MetLysThrAsnProIleValSerThrIleGlyHisPheAspAsnGluIleAspMetPro 358
QY      1087 GGCCTCATGAATACCCAGCATCAAGCACATCCCAATCAAGCCAGAAATACGATGTGG 1146
Db      359 GlyLeuGluThrTyrProGlyValLysArgIleThrIleLysProGlnThrAspArgTyr 378
QY      1147 GAATCCAGATGCGCCAGCT---ATCCTCCTTCTGTCAGGCGCGCTTCTTAACCTT 1203
Db      379 ValPheProGluThrLysAlaGlyIleIleValLeuAlaGluGlyArgLeuMetAsnLeu 398
QY      1204 GGTCTGGTACAGTCAACCATCTTTCGTTATGTCATTCATTCACAAACGACACACTC 1263
Db      399 GlyCysProThrGlyHisProSerPheValMetSerCysSerPheThrAsnGlnValIle 418
QY      1264 GTCAGTCGACTCTACGAAAGAGA-----GGAATCTCGAGAAGAGTTTACACA 1317
Db      419 AlaGlnLeuGluLeuTyrAsnGluLysAlaSerGlyLysTyrGluLysLysValTyrVal 438
QY      1318 CTTCGAGGATCTCGATGAGAGTCGTCGCTCCCTCCACTCGATCTCGATGTCCAC 1377
Db      439 LeuProLysHisLeuAspGluLysValAlaLeuLeuHisLeuGlyLysLeuGlyAlaArg 458
QY      1378 CTTCAGAGCTTACACAGACGAGCTGACTACATCAACCTCCAGTTGAGGCTTCATC 1437
Db      459 LeuThrLysLeuSerLysAspGlnSerAspTyrValSerIleProIleGluGlyProTyr 478
QY      1438 AGTCTGATGCTTACCTTAT 1458
Db      479 LysProProHisTyrArgTyr 485

RESULT 12
AAG41408
ID  AAG41408 standard; Protein; 467 AA.
XX  AAG41408;
AC  AAG41408;
XX  18-OCT-2000 (first entry)
DT  18-OCT-2000 (first entry)
DE  Arabidopsis thaliana protein fragment SEQ ID NO: 51515.
XX  Protein identification; signal transduction pathway; metabolic pathway;
KW  hybridisation assay; genetic mapping; gene expression control; promoter;
KW  termination sequence.
XX  Arabidopsis thaliana.
XX  EP1033405-A2.
XX  06-SEP-2000.
PD  25-FEB-2000; 2000EP-0301439.
XX  25-FEB-2000; 2000EP-0301439.
XX  25-FEB-1999; 99US-0121825.
XX  99US-0123180.
XX  99US-0123548.
XX  99US-0125788.
XX  99US-0126264.
XX  99US-0126785.
XX  99US-0127462.
XX  99US-0128234.
XX  99US-0128714.
XX  99US-0128834.
XX  99US-0129845.
XX  99US-0130077.
XX  99US-0130449.
XX  99US-0130510.
XX  99US-0130891.
XX  99US-0131449.
XX  99US-0132048.
XX  99US-0132407.
XX  99US-0132484.
XX  99US-0132485.
XX  99US-0132486.
XX  99US-0132487.
XX  99US-0132863.
XX  99US-0134256.
XX  99US-0134218.
XX  99US-0134219.
XX  99US-0134221.
XX  99US-0134370.
XX  99US-0134768.
XX  99US-0134941.
XX  99US-0135124.
XX  99US-0135353.
XX  99US-0135629.
XX  99US-0136021.
XX  99US-0136392.
XX  99US-0136782.
XX  99US-0137222.
XX  99US-0137528.
XX  99US-0137502.
XX  99US-0137724.
XX  99US-0138094.
XX  99US-0138540.
XX  99US-0138847.
XX  99US-0139119.
XX  99US-0139452.
XX  99US-0139453.
XX  99US-0139492.
XX  99US-0139454.
XX  99US-0139455.
XX  99US-0139456.
XX  99US-0139457.
XX  99US-0139458.
XX  99US-0139459.
XX  99US-0139460.
XX  99US-0139461.
XX  99US-0139462.
XX  99US-0139463.
XX  99US-0139750.
XX  99US-0139763.
XX  99US-0139817.
XX  99US-0139899.
XX  99US-0140353.
XX  99US-0140354.
XX  99US-0140695.
XX  99US-0140823.
XX  99US-0140991.
XX  99US-0141287.
XX  99US-0141842.
XX  99US-0142154.
XX  99US-0142055.
XX  99US-0142390.
XX  99US-0142803.
XX  99US-0142920.
XX  99US-0142977.
XX  99US-0143542.
```

PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 09-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148319.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.

PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

## Alignment Scores:

Pred. No.: 1,66e-123 Length: 467  
 Score: 1494.00 Matches: 295  
 Percent Similarity: 74.4% Conservative: 55  
 Best Local Similarity: 62.77% Mismatches: 104  
 Query Match: 54.59% Indels: 16  
 DB: 21 Gaps: 4

US-09-759-990A-1 (1-1461) x AAG41408 (1-467)

QY 70 GCGCGTAAGGAACCTACCTGCTGAGAGAGAAATCCAGGCTTATGTTCTTCGTGAG 129  
 Db 7 GlyArgLeuIleGluLeuAlaGluValGluMetProGlyLeuValSerCysValThr 26  
 QY 130 CGTTATTCGCTTCTAAGCCATTGAGGGGTGTGAGATCTCTGGTTCCTCCATCACA 189  
 Db 27 GluPheGlyProSerGlnProLeuLysGlyAlaArgIleThrGlySerLeuHisMetThr 46  
 QY 190 GTCCAGACAGCGCTCCTCATCGAGACACTCACAGCTCTTGGTGTGATGTCAGATGGCT 249  
 Db 47 IleGlnThrAlaValLeuIleGluThrLeuThrAlaLeuGlyAlaGluValArgTrpCys 66  
 QY 250 TCCTGCAACATCTTCTACACAGATACAGCGCTGCTGCTATCGTTGTGGGCCCAACA 309  
 Db 67 SerCysAsnIlePheSerThrGlnAspHisAlaAlaAlaAlaAlaAlaArgAspSerAla 86  
 QY 310 GGCACACAGAGAGCGCGGTATCCAGTCTTCGCTGGAAGGGGCAACACTCCCA 369  
 Db 87 Ala-----ValPheAlaTrpLysGlyGluThrLeuGln 97  
 QY 370 GAATACCTGGGAGAACACATACCGGCTCTCCATGCGCCAGATGGTCAAGGCCACACAGCAG 429  
 Db 98 GluTrpTrpCysThrGluArgAlaLeuAspTrpGlyProGlyGlyProAspLeu 117  
 QY 430 GTTCTGATGATGGTGGTGAATGCTACTCTCTCATCTCCAGGGCTTCGAA-----480  
 Db 118 IleValAspAspGlyGlyAspAlaThrLeuLeuIleHisGluGlyValLysAlaGluGlu 137  
 QY 481 ---TTGGAACAGCGCGTCTCTTCCAGAGCCACAGAGCTGACAACTCGAATACCGC 537

```
Db 138 llePheAlaIysAsnGlyThrPheProAspProThrSerThrAspAsnProGluPheGln 157
QY 538 TGCCTTCTTCTACTCACTCAAGAGGTCTTCAACGACAGAACCACTGCACACAGTT 597
Db 158 lleValLeuSerIleIleIysAspGlyLeuGlnValAspProIysIysThrHisIysMet 177
QY 598 GCTCCGGCATGACCGTCTTCCGACAGACACACACAGGTGTCCACGGCTCTACAG 657
Db 178 LysGluArgLeuValGlyValSerGluGluThrThrGlyValIysArgLeuTyrGln 197
QY 658 CTCGAGAGAGAGGCGCAAACTCTCTTCCAGCCATCAACCTCAACGACGCTGTTCACAAAG 717
Db 198 MetGluGluThrGlyAlaLeuLeuPheProAlaIleAsnValAsnAspSerValThrLys 217
QY 718 TCCAAAGTTCGATTAACATCTACGGCTGCGCCACTCCCTTATCATGATGATCAACCGTCT 777
Db 218 SerIysPheAspAsnLeuTyrGlyCysArgHisSerLeuProAspGlyLeuMetArgAla 237
QY 778 TCCGATCTCATGATCGCGGCAAGACACTCTCTGTCATGGTTCAGCGATGTCGGAAG 837
Db 238 ThrAspValMetIleAlaGlyLysValAlaValIleCysGlyTyrGlyAspValGlyLys 257
QY 838 GCGTCGCGCTCAATCTCGCTCGCCGCAAGCGCTCGCTTATCATCACAAGAACTCGACCCA 897
Db 258 GlyCysAlaAlaAlaMetIysThrAlaGlyAlaArgValIleValThrGluIleAspPro 277
QY 898 ATCTGCGCTCTCAGGCTGCCATGGAAGGCTACAGGTCCGCGCATCGAGAAAGTCGTC 957
Db 278 IleCysAlaLeuGlnAlaLeuMetGluGlyLeuGlnValLeuThrLeuGluAspValVal 297
QY 958 AAGATCTCATGATCTTCTTACATGACACAGAACTGGGATATCATCTCTGTTGACATG 1017
Db 298 SerGluAlaAspIlePheCysThrThrThrGlyAsnLysAspIleIleMetValAspHis 317
QY 1018 ATGCCCCAGATGAAGGCTATTGTCGTAACATCGGCCACTTCGATAACGAAAT 1077
Db 318 MetArgLysMetIysAsnAsnAlaIleValCysAsnIleGlyHisPheAspAsnGluIle 337
QY 1078 GATACAGATGGCTGATGAATACCCAGGATCAGACATCCCAATCAACGCCAGATAC 1137
Db 338 AspMetLeuGlyLeuGluThrTyrProGlyValIysArgIleThrIleLysProGlnThr 357
QY 1138 GACATGNGGAATTCACGATCGCCAGCT---ATCTCTCTTCTGCTGAGGCGCGCTT 1194
Db 358 AspArgIrpValPheProAspThrAsnSerGlyIleIleValLeuAlaGluGlyArgLeu 377
QY 1195 CTAAACCTTGGCTGCGCTACAGGTCAACCATCTTTCGTTATGTCATGTCATTCAACAAC 1254
Db 378 MetAsnLeuGlyCysAlaThrGlyHisProSerPheValMetSerCysSerPheThrAsn 397
QY 1255 CAGACATCTGCTCAGCTCAGCTTACGAAAGAGA-----GGAAATCTCGAGAGAAG 1308
Db 398 GlnValIleAlaGlnLeuGluLeuThrPAsnGluLysSerSerGlyLysTyrGluLysLys 417
QY 1309 GTTTACACATTCGGAAGCATCTCCATGAAGAAGTGTGCTCGCTCCACTCGGATCTCTC 1368
Db 418 ValTyrValLeuProIysHisLeuAspGluLysValAlaAlaLeuHisLeuGlyLysLeu 437
QY 1369 GATGTCACCTTCAAAAGCTTACACAGACAGCTGACTCATCATCAACAGTTCACAGTTGAG 1428
Db 438 GlyAlaArgLeuThrLysLeuThrLysAspGlnSerAspTyrValSerIleProValGlu 457
QY 1429 GGTCTTACAGTCTGATCGTTACCGTTAT 1458
Db 458 GlyProTyrIysProValHisTyrArgTyr 467
RESULT 13
AAW01458
ID AAW01458 standard; Protein; 485 AA.
XX
AC
XX
AAW01458;
```

```
DT 22-FEB-1997 (first entry)
XX
DE Asparagus S-adenosyl-L-homocysteine hydrolase.
XX
KW S-adenosyl-L-homocysteine hydrolase; SHH; promoter; monocot; dicot;
transgenic plant; disease resistance; pathogen resistance.
XX
OS Asparagus officinalis.
XX
FH Key Location/Qualifiers
FT Region 150..190
FT /note= "region found in SHH of photosynthetic
FT species"
FT Binding-site 263..294
FT /label= NAD+_binding_site
XX
PN W09632488-A1.
XX
PD 17-OCT-1996.
XX
PF 10-APR-1996; 96WO-GB00882.
XX
PR 10-APR-1995; 95GB-0007381.
XX
PA (ZENE ) ZENECA LTD.
XX
PI Draper J, Greenland AJ, Skipsey M, Warner S;
XX
DR WPI; 1996-477138/47.
DR N-PSDB; AAT44513.
XX
PT S-adenosyl-L-homocysteine hydrolase promoter - used for driving
expression of effector genes, such as pathogen resistance genes, in
transgenic plants
XX
PS Example 1; Fig 1; 57bp; English.
XX
CC Asparagus S-adenosyl-L-homocysteine hydrolase (SHH) (AAW01458)
catalyses the reversible hydrolysis of S-adenosyl-L-homocysteine
to adenosine and homocysteine. It contains an extra stretch of
amino acid residues (positions 150-190) previously found in other
photosynthetic species, parsley and Rhodobacter capsulatus, but not
in SHHs from non-photosynthetic species. A cDNA sequence (AAT44513)
coding for the asparagus was used to identify the Arabidopsis
thaliana SHH gene (AAT44515) and promoter (AAT44514), useful for
expression of effector genes in transgenic plants.
XX
SQ Sequence 485 AA;
Alignment Scores:
Pred. No.: 3,82e-123 Length: 485
Score: 1490.00 Matches: 294
Percent Similarity: 72.82% Conservative: 57
Best Local Similarity: 61.00% Mismatches: 115
Query Match: 54.44% Indels: 16
DB: 17 Gaps: 4
US-09-759-990A-1 (1-1461) x AAW01458 (1-485)
QY 34 GAGTACAGAAATTCGCGCATCAATCACTCCATGTTCTCGCGCGTAAGAACTTACGCTTGCT 93
Db 13 GluTyrIysValIysAspMetSerGlnAlaAspPheGlyArgLeuGluIleGluLeuAla 32
QY 94 GAGAAAGGAAATGCCAGGTCTTATGGTTCTTTCGTGACGGTTATTCGGCTTCTTAAGCCATTG 153
Db 33 GluValGluMetProGlyLeuMetAlaCysArgAlaGluPheGlyProAlaGlnProPhe 52
QY 154 AAGGTTCTCAGATCTCTGGTTCCTCCATCAGACTCCAGACGGCTCTCATCGAG 213
Db 53 LysGlyAlaLysIleThrGlySerLeuHisMetThrIleGlnThrAlaValLeuIleGlu 72
QY 214 AACTCAGACTCTTGGTGTGATGTGATGGCTTCCTGCAACATCTTCTCTACACAA 273
```





PR 07-JUN-1999;	99US-0137724.	PR 06-AUG-1999;	99US-0147416.
PR 08-JUN-1999;	99US-0138094.	PR 09-AUG-1999;	99US-0147493.
PR 10-JUN-1999;	99US-0138540.	PR 09-AUG-1999;	99US-0147935.
PR 10-JUN-1999;	99US-0138847.	PR 10-AUG-1999;	99US-0148171.
PR 14-JUN-1999;	99US-0139119.	PR 11-AUG-1999;	99US-0148319.
PR 16-JUN-1999;	99US-0139452.	PR 12-AUG-1999;	99US-0148341.
PR 16-JUN-1999;	99US-0139453.	PR 13-AUG-1999;	99US-0148565.
PR 17-JUN-1999;	99US-0139492.	PR 13-AUG-1999;	99US-0148684.
PR 18-JUN-1999;	99US-0139454.	PR 16-AUG-1999;	99US-0149368.
PR 18-JUN-1999;	99US-0139455.	PR 17-AUG-1999;	99US-0149175.
PR 18-JUN-1999;	99US-0139456.	PR 18-AUG-1999;	99US-0149426.
PR 18-JUN-1999;	99US-0139457.	PR 20-AUG-1999;	99US-0149722.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149723.
PR 18-JUN-1999;	99US-0139459.	PR 20-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139460.	PR 23-AUG-1999;	99US-0149902.
PR 18-JUN-1999;	99US-0139461.	PR 23-AUG-1999;	99US-0149930.
PR 18-JUN-1999;	99US-0139462.	PR 25-AUG-1999;	99US-0150566.
PR 18-JUN-1999;	99US-0139463.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139750.	PR 27-AUG-1999;	99US-0151065.
PR 18-JUN-1999;	99US-0139763.	PR 27-AUG-1999;	99US-0151066.
PR 21-JUN-1999;	99US-0139817.	PR 27-AUG-1999;	99US-0151080.
PR 22-JUN-1999;	99US-0139899.	PR 30-AUG-1999;	99US-0151303.
PR 23-JUN-1999;	99US-0140353.	PR 31-AUG-1999;	99US-0151438.
PR 23-JUN-1999;	99US-0140354.	PR 01-SEP-1999;	99US-0151930.
PR 23-JUN-1999;	99US-0140695.	PR 07-SEP-1999;	99US-0152363.
PR 28-JUN-1999;	99US-0140823.	PR 10-SEP-1999;	99US-0153070.
PR 29-JUN-1999;	99US-0140991.	PR 13-SEP-1999;	99US-0153758.
PR 30-JUN-1999;	99US-0141287.	PR 15-SEP-1999;	99US-0154018.
PR 01-JUL-1999;	99US-0141842.	PR 16-SEP-1999;	99US-0154039.
PR 01-JUL-1999;	99US-0142154.	PR 20-SEP-1999;	99US-0154779.
PR 02-JUL-1999;	99US-0142055.	PR 22-SEP-1999;	99US-0155139.
PR 08-JUL-1999;	99US-0142390.	PR 23-SEP-1999;	99US-0155486.
PR 08-JUL-1999;	99US-0142803.	PR 24-SEP-1999;	99US-0155659.
PR 09-JUL-1999;	99US-0142920.	PR 28-SEP-1999;	99US-0156458.
PR 12-JUL-1999;	99US-0142977.	PR 29-SEP-1999;	99US-0156596.
PR 13-JUL-1999;	99US-0143542.	PR 04-OCT-1999;	99US-0157117.
PR 14-JUL-1999;	99US-0143624.	PR 05-OCT-1999;	99US-0157753.
PR 15-JUL-1999;	99US-0144005.	PR 06-OCT-1999;	99US-0157865.
PR 16-JUL-1999;	99US-0144085.	PR 07-OCT-1999;	99US-0158029.
PR 16-JUL-1999;	99US-0144086.	PR 08-OCT-1999;	99US-0158232.
PR 19-JUL-1999;	99US-0144325.	PR 12-OCT-1999;	99US-0158369.
PR 19-JUL-1999;	99US-0144331.	PR 13-OCT-1999;	99US-0159293.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144333.	PR 13-OCT-1999;	99US-0159295.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159329.
PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159330.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159637.
PR 20-JUL-1999;	99US-0144894.	PR 14-OCT-1999;	99US-0159638.
PR 21-JUL-1999;	99US-0144814.	PR 18-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0145036.	PR 21-OCT-1999;	99US-0160741.
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160767.
PR 22-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160768.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160770.
PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160814.
PR 22-JUL-1999;	99US-0145192.	PR 21-OCT-1999;	99US-0160815.
PR 23-JUL-1999;	99US-0145145.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160981.
PR 23-JUL-1999;	99US-0145224.	PR 22-OCT-1999;	99US-0160989.
PR 26-JUL-1999;	99US-0145276.	PR 25-OCT-1999;	99US-0161404.
PR 27-JUL-1999;	99US-0145913.	PR 25-OCT-1999;	99US-0161405.
PR 27-JUL-1999;	99US-0145918.	PR 25-OCT-1999;	99US-0161406.
PR 27-JUL-1999;	99US-0145919.	PR 26-OCT-1999;	99US-0161359.
PR 28-JUL-1999;	99US-0145921.	PR 26-OCT-1999;	99US-0161360.
PR 02-AUG-1999;	99US-0146386.	PR 26-OCT-1999;	99US-0161361.
PR 02-AUG-1999;	99US-0146388.	PR 28-OCT-1999;	99US-0161920.
PR 02-AUG-1999;	99US-0146389.	PR 28-OCT-1999;	99US-0161992.
PR 03-AUG-1999;	99US-0147038.	PR 28-OCT-1999;	99US-0161993.
PR 04-AUG-1999;	99US-0147204.	PR 29-OCT-1999;	99US-0162142.
PR 05-AUG-1999;	99US-0147302.		
PR 05-AUG-1999;	99US-0147192.		
PR 05-AUG-1999;	99US-0147260.		
PR 06-AUG-1999;	99US-0147303.		

Alignment Scores:  
Pred. No.:  
Score:

1.18e-121  
1473.00

Length:  
Matches:

450  
290

Db 312 AsnIleGlyHisPheAspAsnGluIleAspMetLeuGlyLeuGluThrTyrProGlyVal 331  
 Qy 1111 AAGCACATCCCAATCAAGCCAGAATACGACATCTGGGAATCCAGATGCCAC--GCT 1167  
 Db 332 LysArgIleThrIleLysProGlnThrAspArgTtpValPheProGluThrAsnThrGly 351  
 Qy 1168 ATCTCTCTTTCCTGAGGGCGGCTCTTAACCTTGGCTGGCTACAGTCAACCATCT 1227  
 Db 352 IleIleValLeuAlaGluGlyArgLeuMetAsnLeuGlyCysAlaThrGlyHisProSer 371  
 Qy 1238 TTCCTTATGTCAATGTCAATCCACAAACACACACTCGCTCGAGCTCGACCTACGAAAG 1287  
 Db 372 PheValMetSerCysSerPheThrAsnGlnValIleAlaGlnLeuGluLeuTyrIysGlu 391  
 Qy 1288 AGA-----GGAAATCTCGAGAGAGAGTTTACACACTTCGGAAGCATCTCGATGAAGA 1341  
 Db 392 LysSerSerGlyLysTyrGluLysLysValTyrValLeuProLysHisLeuAspGluLys 411  
 Qy 1342 GTCTCTGGCTCCACTCCGATCTCTCGATGTCACCTTACAAAGCTTACACAGCTTACAGAGCAG 1401  
 Db 412 ValAlaAlaLeuHisLeuGlyLysLeuGlyAlaLysLeuThrLysLeuThrIysSerGln 431  
 Qy 1402 GCTGACTACATCAAGTTCCAGTTGAGGTCCTTCAAGTCTGATCGCTTACCGCTTAT 1458  
 Db 432 AlaAspTyrIleSerValProIleGluGlyProTyrLysProAlaHisTyrArgTyr 450  
 RESULT 15  
 AAG22957  
 ID AAG22957 standard; Protein; 467 AA.  
 XX AC AAG22957;  
 XX AC  
 DT 17-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 26083.  
 XX  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX  
 OS Arabidopsis thaliana.  
 PN EP1033405-A2.  
 XX  
 XX 06-SEP-2000.  
 XX  
 XX 25-FEB-2000; 2000EP-0301439.  
 XX  
 PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0134445.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR

PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135153.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137582.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 18-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150684.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160880.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: October 15, 2003, 14:43:19 ; Search time 29 Seconds  
(without alignments)

4263.184 Million cell updates/sec

Title: US-09-759-990A-1

Perfect score: 2737  
Sequence: 1 atggcttgcaatcacctac.....ctgatgcttaccgttattaa 1461

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 657434

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer\_n2p.model -DEV=xl  
-Q=/cgn2\_1/USPTO\_spool/US0975990/runat 15102003 115154 2410/app query.fasta\_1.1607  
-DB=Issued Patents AA -OPMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0975990 -CGN 1\_1\_38 -runat 15102003 115154 2410 -NCPV=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA.\*

1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/ECTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1490	54.4	485	3	US-08-930-894-2
2	1278	46.7	432	2	US-08-896-005-5
3	1255	45.9	432	2	US-08-896-005-4
4	1255	45.9	432	4	US-09-347-878-1
5	1081.5	33.5	504	2	US-08-896-005-3
6	1078	39.4	463	4	US-09-328-352-4600
7	1061	38.8	502	4	US-09-252-991A-23501
8	1033	37.7	500	2	US-08-896-005-1
9	421.5	16.1	215	4	US-09-252-991A-23397
10	380.5	14.5	494	4	US-09-252-991A-23386
11	351	12.8	140	3	US-08-930-894-4
12	335.5	12.3	139	3	US-08-930-894-6

13	323.5	11.8	138	3	US-08-930-894-5	Sequence 5, Appli
14	307.5	11.2	139	3	US-08-930-894-7	Sequence 7, Appli
15	287	10.5	97	1	US-08-204-740-7	Sequence 7, Appli
16	287	10.5	97	3	US-08-081-167A-7	Sequence 7, Appli
17	287	10.5	97	3	US-09-081-395-7	Sequence 7, Appli
18	287	10.5	97	3	US-09-416-833-7	Sequence 7, Appli
19	287	10.5	97	5	PCT-US95-02521-7	Sequence 9, Appli
20	271	9.9	95	1	US-08-204-740-9	Sequence 9, Appli
21	271	9.9	95	3	US-09-081-167A-9	Sequence 9, Appli
22	271	9.9	95	3	US-09-081-395-9	Sequence 9, Appli
23	271	9.9	95	3	US-09-081-833-9	Sequence 9, Appli
24	271	9.9	95	5	PCT-US95-02521-9	Sequence 9, Appli
25	211	7.7	2137	4	US-09-134-001C-4463	Sequence 8, Appli
26	188.5	6.9	99	3	US-08-930-894-8	Sequence 4463, Ap
27	172.5	6.3	421	4	US-09-252-991A-22123	Sequence 22123, A
28	172.5	6.3	1481	2	US-08-616-844-40	Sequence 40, Appl
29	172.5	6.3	1481	2	US-08-599-654-40	Sequence 40, Appl
30	172.5	6.3	1481	3	US-08-944-868A-40	Sequence 40, Appl
31	172.5	6.3	1481	3	US-08-944-423A-40	Sequence 40, Appl
32	172.5	6.3	1481	3	US-08-944-496-40	Sequence 40, Appl
33	156	5.7	714	4	US-09-252-991A-31209	Sequence 31209, A
34	155	5.7	997	4	US-09-252-991A-30799	Sequence 30799, A
35	155	5.7	1053	4	US-09-252-991A-26140	Sequence 26140, A
36	154.5	5.6	822	4	US-09-252-991A-21920	Sequence 21920, A
37	154	5.6	618	4	US-09-252-991A-28358	Sequence 28358, A
38	153	5.6	430	3	US-09-156-836B-2	Sequence 2, Appli
39	152.5	5.6	618	4	US-09-252-991A-23696	Sequence 23696, A
40	150.5	5.5	551	4	US-09-252-991A-23555	Sequence 23555, A
41	150	5.5	430	3	US-08-997-897-2	Sequence 2, Appli
42	149	5.4	493	4	US-09-252-991A-16925	Sequence 16925, A
43	148.5	5.4	512	4	US-09-252-991A-31457	Sequence 31457, A
44	146.5	5.4	373	4	US-09-252-991A-29008	Sequence 29008, A
45	145	5.5	102	4	US-09-252-991A-23296	Sequence 23296, A

#### ALIGNMENTS

RESULT 1  
US-08-930-894-2  
; Sequence 2, Application US/08930894  
; Patent No. 6037524  
; GENERAL INFORMATION:  
; APPLICANT: GREENLAND, Andrew James  
; APPLICANT: DRAPER, John  
; APPLICANT: SKIPSEY, Marc  
; APPLICANT: WARNER, Simon  
; TITLE OF INVENTION: S-ADENOSYL-L-HOMOCYSTEIN HYDROLASE PROMOTER  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pillsbury Madison & Sutro  
; STREET: 1100 New York Avenue, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3918  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: MS word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/930,894  
; FILING DATE: 09-OCT-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB96/00882  
; FILING DATE: 10-APR-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9507381.3  
; FILING DATE: 10-APR-1995  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 485 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-930-894-2

## Alignment Scores:

Pred. No.: 1.53e-127 Length: 485  
Score: 1490.00 Matches: 294  
Percent Similarity: 72.92% Conservative: 57  
Best Local Similarity: 61.00% Mismatches: 115  
Query Match: 54.44% Indels: 16  
DB: 3 Gaps: 4

US-09-759-990A-1 (1-1461) x US-08-930-894-2 (1-485)

```

QY 34 GAGTACAGATTGGCGACATCAACCTCCATGTTCTCGGCGTAAGAACTTACCTGCT 93
DB 13 GluTyrLysValLysAspMetSerGlnAlaaspPheGlyArgLeuGluLeuAla 32
QY 94 GAGAAGAAATGCCAGGTCTTATGGTCTTCCTGAGCGTATTCCTGCTTCAAGCAATG 153
DB 33 GluValGluMetProGlyLeuMetAlaCysArgAlaGluPheGlyProAlaGlnProPhe 52
QY 154 AAGGGTGTGAGATCTGCTGCTCCATGACATGACATCCAGACAGCGCTCATCGAG 213
DB 53 LysGlyAlaLysIleThrGlySerLeuHisMetThrIleGlnThrAlaValLeuIleGlu 72
QY 214 ACATCCAGCTCTTGGTGTGATGTCAGATGGGCTTCCTGCAACATCTTCTACACAA 273
DB 73 ThrLeuThrAlaLeuGlyProGluValArgTrpCysSerCysAsnIlePheSerThrGln 92
QY 274 GATACAGCGCTGCTGCTATGCTGTGTCGGCCCAACAGGCACACAGAGAGCCAGCGGT 333
DB 93 AspHisAlaAlaAlaAlaIleAlaArgAspSerAlaSer-----105
QY 334 ATCCAGTCTTCGCTGGAAGGGCGAAACACTCCAGAACTACGCGGAGAACACATACGCG 393
DB 106 -----ValPheAlaIleTrpLysGlyGluThrLeuGlnGluTyrTrpCysThrGluArg 123
QY 394 GCTCTCAGATGCCAGATGTGTCAGGCCACAGCAGGTGTGTCAGATGATGGTGTGATGCT 453
DB 124 AlaLeuAspTrpGlyProGlyGlyGlyProAspLeuIleValAspAspGlyGlyAspThr 143
QY 454 ACATCTCTCATCTCCAGGCGTTC-----GAATTGGAACACGCGTCTGTT 501
DB 144 ThrLeuLeuIleHisGluGlyValLysAlaGluGluGluTyrGluLysThrGlyLysMet 163
QY 502 CGAGGCCAACAGAGCTGACAACTCGAATACGCTGCGTCTTCTGCTACACTCAAGCAG 561
DB 164 ProAspProAlaSerThrAspAsnAlaGluPheGlnIleValLeuThrIleIleArgAsp 183
QY 562 GTCTTCAACCAAGACAGAACCACTGSCACACAGTGTGCGCGCATGAACGCTGTTCC 621
DB 184 GlyLeuLysValAspProThrLysTyrArgLysMetLysAspArgIleValGlyValSer 203
QY 622 GAAGAGAACACAGAGTGTCCACCGCTCTACAGCTCGAGAGAGGAGGCGCAACTCTC 681
DB 204 GluGluThrThrGlyValLysArgLeuTyrGlnMetGlnAlaAsnAsnSerLeuLeu 223
QY 682 TTCCGAGCATCAAGCTCAAGCTGTITACAAAGTCCAAAGTTCATGATCGCGGCAAG 741
DB 224 PheProAlaIleAsnValAsnAspSerValThrLysSerLysPheAspAsnLeuTyrGly 243
QY 742 TGGCGGCACTCTCTATCGATGGTATCAACCGTGTCTCCGATGTCATGATCGCGGCAAG 801
DB 244 CysArgHisSerLeuProAspGlyLeuMetArgAlaThrAspValMetIleAlaGlyLys 263
QY 802 ACAGCTCTGCTCATGGTATCGCGATGTCGCGAGCGCTCGCTCAATCCCTCGGTGGC 861
DB 264 ValAlaValValCysGlyTyrGlyAspValGlyGluGlyCysAlaAlaLeuLysGln 283
QY 862 CAAGGGCGCTGCTTATCATCAGAGACTCGACCAATCTCGCTCTCCAGGCTCGCATG 921

```

```

DB 284 AlaGlyAlaArgValIleValThrGluIleAspProIleCysAlaLeuGlnAlaLeuMet 303
QY 922 GAAGCTTACAGAGTCCGCGCATCGAGAAAGTCTCAGAGATGTCGATATCTCGTTACA 981
DB 304 GluGlyLeuGlnValIleThrLeuGluAspValValSerGluAlaAspIlePheValThr 323
QY 982 TSCACAGAAACTGGCATATCATCTCTGTTGACATGATGGCCAGATGAAGATAAGGT 1041
DB 324 ThrThrGlyAsnLysAspIleIleMetLeuAspHisMetArgLysMetLysAsnAla 343
QY 1042 ATTGTGCTTAACATCGGCACTTCGATAACGAAATGTATACAGATGGCTCATCAATAC 1101
DB 344 IleValCysAsnIleGlyHisPheAspAsnGluIleAspMetLeuGlyLeuGluThrThr 363
QY 1102 CCAGGCATCAAGCACATCCCAATCAAGCCAGATACGACATGTCGGAATCCCAAGATGC 1161
DB 364 ProGlyIleLysArgIleThrIleLysProGlnThrAspArgTrpValPheProGluThr 383
QY 1162 CAC---GCTATCTCTCTTCTGAGGGCGGCTTCTTAACCTTGGCTCGCTACAGT 1218
DB 384 AsnThrGlyIleIleValIleAlaGluGlyArgLeuMetAsnLeuGlyCysAlaThrGly 403
QY 1219 CACCATCTTTCGTTATGTCAATGTCAATTCACAAACAGACACTCGCTCAGCTCGACCTC 1278
DB 404 HisProSerPheValMetSerCysSerPheThrAsnGlnValIleAlaGlnLeuGluLeu 423
QY 1279 TACGAAAGAGA-----GGAAATCTCGAGAGAGGTTTACACTTCGGAAGCATCTC 1332
DB 424 TrpAsnGluLysAlaSerGlyLysTyrGluLysLysValTyrValLeuProLysHisLeu 443
QY 1333 GATCAGAAAGTCTGCGCTCCACTCGGATCTCTCGATGTCCACCTTACAAAGCTTACA 1392
DB 444 AspGlnLysValAlaAlaLeuHisLeuGlyLysLeuGlyAlaLysLeuThrLysLeuSer 463
QY 1393 CAGAAGCAGCTGACTATACATCAAGTTCAGTTCAGGTTCCTTCAAGTCTGATGCTTAC 1452
DB 464 ProSerGlnAlaAspTyrIleSerValProIleGluGlyProTyrLysProProHisTyr 483
QY 1453 GGTAT 1458
DB 484 ArgTyr 485

```

## RESULT 2

US-08-896-005-5  
Sequence 5, Application US/08896005  
Patent No. 5854023  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Corley, Neil C.  
APPLICANT: Lal, Preeti  
APPLICANT: Shah, Purvi  
TITLE OF INVENTION: S-ADENOSYL-L-HOMOCYSTEINE HYDROLASE  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESS: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/896,005  
FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749  
 REFERENCE/DOCKET NUMBER: PF-0337 US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-855-0555  
 TELEFAX: 415-845-4166  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 432 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 LIBRARY: GenBank  
 CLONE: 904132  
 US-08-896-005-5

## Alignment Scores:

Pred. No.: 3,33e-108 Length: 432  
 Score: 1278.00 Matches: 270  
 Percent Similarity: 66.74% Conservative: 49  
 Best Local Similarity: 56.49% Mismatches: 103  
 Query Match: 46.69% Indels: 56  
 DB: 2 Gaps: 5

US-09-759-990A-1 (1-1461) x US-08-896-005-5 (1-432)

```

QY 37 TACAGATTGCGCATACACTCCATGTTCTCGCCGTAAGCAACTACCTTGTCTGAG 96
D 7 TyrLysValAlaAspIleGlyLeuAlaAlaTrpGlyArgLysAlaLeuAspIleAlaGlu 26
QY 97 AGGAAATGCGAGCTTATGTTCTTCTGAGCGTTATTCGCTTCTAAGCCATCAAG 156
D 27 AsnGluMetProGlyLeuMetArgMetArgGluMetTyrSerAlaSerLysProLeuLys 46
QY 157 GGTGTGAGAACTCTGCTTCCATGACAGTCCAGACAGCCGCTCTCATCGAGACA 216
D 47 GlyAlaAlaGlyLeuAlaGlyCysLeuHisMetThrValGluThrAlaValLeuIleuThr 66
QY 217 CTCACAGCTCTGCTGATGTCAGATGGCTTCTCGCAACATCTTCTTACACAGAT 276
D 67 LeuValAlaLeuGlyAlaGluValArgTrpSerCysAsnIlePheSerThrGlnAsp 86
QY 277 ACAGCGGTGCTGCTATGCTTGTGCGCCCAACAGAGCACACAGAGAGCCGCGTATC 336
D 87 HisAlaAlaAlaAlaIleAla-----LysAlaGlyIle 97
QY 337 CAGTCTTCGCTGGAAGGCGGAAACATCCAGAAATCTGGAGAACACATACCGCGCT 396
D 98 ProValPheAlaTrpLysGlyGluThrAspGluGluTyrLeuTrpCysIleGluGlnThr 117
QY 397 CTCACATGCGCCAGATGTCACGCCCCACAGCAGGTTGTCATGATGTCGTGATGCTACA 456
D 118 LeuHisPheLysAspGlyPro----LeuAsnMetIleLeuAspAspGlyGlyAspLeuThr 136
QY 457 CTCCTCATCTCCAAAGGCTTCGAATTCGAATTCGAACAGCCGCTGCTGTCACAGCAACAGAA 516
D 137 AsnLeuIle-----139
QY 517 GCTGACACCTCGATACCGCTCGTTCCTTGTCTACACTCAGCAGGCTCTTCACACAGAC 576
D 139 -----139
QY 577 AAGAACCACTGCGCACACA-----GTTGCTGCGCGCATGAACGCTGTTTCCGAA 624
D 140 -----HisThrLysTyrProGlnLeuLeuSerGlyIleArgGlyIleSerGlu 155
QY 625 GAGACAAACAGGTGTCACCGCTCTTACAGCTCGAAGAGGAGGCGCAAACTCTCTCTTC 684
D 156 GluThrThrThrGlyValHisAsnLeuTyrLysMetSerAsnGlyIleLeuAsnVal 175
QY 685 CCGAGCATCAAGCTCAAGCGCTGTACAAAGTCCAGATTCGATACATCTACCGCTGC 744
D 176 ProAlaIleAsnValAsnAspSerValThrLysSerLysPheAspAsnLeuTyrGlyCys 195

```

```

QY 745 CGCACTCCCTTATCGATGGTATCAACCGCTGCTCCGATGTCATGATCGCGCGCAAGACA 804
D 196 ArgGluSerLeuIleAspGlyIleLysArgAlaThrAspValMetIleAlaGlyLysVal 215
QY 805 GCTCTGCTCATGGTTACGGCGATGTCGGCAAGGCTCGCTCAATCCCTCCGTCGCCAA 864
D 216 AlaValValAlaGlyTyrGlyAspValGlyLysGlyCysAlaGlnAlaLeuArgGlyPhe 235
QY 865 GCGCTCGCGTTATCATCACAGAACTCGACCAATCTCGCTCTCCAGGTGCTCAATG 924
D 236 GlyAlaArgValIleIleThrGluIleAspProIleAsnAlaLeuGlnAlaAlaMetGlu 255
QY 925 GGTACACAGTCCGCGCATCGAGGAAGTCGTCAAGGATGTCGATATCTTCTGTTATCATGC 984
D 256 GlyTyrGluValThrThrMetAspGluAlaCysLysGluGlyAsnIlePheValThrThr 275
QY 985 ACAGGAACCTGCGATATCATCTCTGTTGACATGATGGCCAGATGAGGATGAAGCTATT 1044
D 276 ThrGlyCysValAspIleIleLeuGlyArgHisPheGluGlnMetLysAspAlaIle 295
QY 1045 GTCGTATACATCGGCCACTTCGATAACGAATTTGATACAGATGGCCTCATGAAATACCCA 1104
D 296 ValCysAsnIleGlyHisPheAspValGluIleAspValLysTrpLeuAsnGlu---Asn 314
QY 1105 GGCATCAAGCATCCCAATCAAGCCAGATAGCATGTGGAAATCCCAAGATGCCAC 1164
D 315 AlaValGluLysValAsnIleLysProGlnValAspArgTyrTrpLeuLysAsnGlyArg 334
QY 1165 GCTATCTCTCTTCTGTCGAGCGCGCTTCTTAACCTTGGCTCGCTACAGTCAACCCA 1224
D 335 ArgIleIleLeuAlaGluGlyArgLeuValAsnLeuGlyCysAlaMetGlyHisPro 354
QY 1225 TCTTTGCTATGTCATCTCAATTCACAACACAGACATCGCTCGCTCAGCTCGACCTCTACGAA 1284
D 355 SerPheValMetSerAsnSerPheThrAsnGlnValMetAlaGlnIleGluLeuThrThr 374
QY 1285 AAGAGGAGAAATCTCGAGAAGAGTTTACACACTTCGGAAGCATCTCGATGAAGATC 1344
D 375 HisProAspLysTyrProValGlyValHisPheLeuProLysLysLeuAspGluAlaVal 394
QY 1345 GCTCGCTCCACTCCGATCTCTCGATGTCACCTTACAAAGCTTACACAGAGAGGCT 1404
D 395 AlaGluAlaHisLeuGlyLysLeuAsnValLysLeuThrLysLeuThrGluLysGlnAla 414
QY 1405 GACTACATCAAGTTCAGGTGAGGCTCTTACAGTCTGATGCTTACCGTTAT 1458
D 415 GlnTyrLeuGlyMetProIleAsnGlyProPheLysProAspHisTyrArgTyr 432

```

## RESULT 3

US-08-896-005-4  
 Sequence 4, Application US/08896005  
 Patent No. 5854023  
 GENERAL INFORMATION:  
 APPLICANT: Hillman, Jennifer L.  
 APPLICANT: Corley, Neil C.  
 APPLICANT: Lal, Preeti  
 APPLICANT: Shah, Purvi  
 TITLE OF INVENTION: S-ADENOSYL-L-HOMOCYSTEINE HYDROLASE  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 STREET: 3174 Porter Drive  
 CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94304  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:



```

, APPLICATION NUMBER: US/08/896,005
, FILING DATE: Filed Herewith
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER:
, FILING DATE:
, ATTORNEY/AGENT INFORMATION:
, NAME: Billings, Lucy J.
, REGISTRATION NUMBER: 36,749
, REFERENCE/DOCKET NUMBER: PF-0337
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: 415-855-0555
, TELEFAX: 415-845-4166
, INFORMATION FOR SEQ ID NO: 4:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 432 amino acids
, TYPE: amino acid
, STRANDEDNESS: single
, TOPOLOGY: linear
, IMMEDIATE SOURCE:
, LIBRARY: GenBank
, CLONE: 178277
, US-08-896-005-4

```

Alignment Scores:

Argument Scores:					
Pred. No.:	4.2e-106	Length:	432		
Score:	1255.00	Matches:	265		
Percent Similarity:	66.32%	Conservative:	52		
Best Local Similarity:	55.44%	Mismatches:	105		
Query Match:	45.85%	Indels:	56		
DB:	2	Gaps:	5		

US-09-759-990A-1 (1-1461) x US-08-896-005-4 (1-432)

37	TACAGAA	NTCGGACATCAAC	CACTCCATG	TCTTCGGCCG	TAAAGAACTTAC	CCCTTGTGTGAG	96
QY							
Db	7	TyrIysValAlaAspIleGlyLeuAlaAla	TrpGlyArgLysAlaLeuAspIleAlaGlu	26			
QY	97	AAGGAAATGCCAGGTCCTTAATGGTCTTCGCTGAGGGTTATCCGCTTCTAAAGCATTAAG	156				
Db	27	AsnGluMetProGlyLeuMetArgMetArgLysTrpSerAlaSerLysProLeuLys	46				
QY	157	GGTGTCAGAATCTCTGGTTCCTCCACATGACACAGTCCACAGCCGCTCCTCATCGAGACA	216				
Db	47	GlyAlaArgIleAlaGlyCysLeuHisMetThrValGluThrAlaValLeuIleGluThr	66				
QY	217	CTCCAGGCTCTTGGTGTGATGTCAGATGGGCTTCTCGCAACATCTTCTCTACACAGAT	276				
Db	67	LeuValThrLeuGlyAlaGluValGlnTrpSerSerCysAsnIlePheSerThrGlnAsn	86				
QY	277	ACAGCCGCTGCTCATCTGCTTCGGCCACACAGGCACACAGAACCCAGCCGGTATC	336				
Db	87	HisAlaAlaAlaIleAla	-----LysAlaGlyIle	97			
QY	337	CCAGTCTTCCGCTGGAAGGCGCAACACTCCACAGAACTATGGGGAACACATACCCGGCT	396				
Db	98	ProValTyrAlaAlaTrpLysGlyIuThrAspGluIuTyrLeuTrpCysIleGluGlnThr	117				
QY	397	CTCCATGCGCCAGATGGTCAGGCCCAACAGCAGGTTCGTGATGTGGTGATGCTACA	456				
Db	118	LeuTyrPheLysAspGlyPro--LeuAsnMetIleLeuAspGlyGlyAspLeuThr	136				
QY	457	CTCTCATCTCCAGGGCTTCGAATTCGAACACCGGCTGCTTCCAGAGCCACACAGAA	516				
Db	137	AsnLeuIle	-----	139			
QY	517	GCTGCACACCTCGAATACCGCTCGCTTCTTGCTACACTCAAGCAGGTCTTCAACCAAGAC	576				
Db	139		-----	139			
QY	577	AAGAACCCTGGCACACA	-----GTTCTGCTCGGATGATGACGGTGTTCGGAA	624			
Db	140	-----HisThrLysTyrProGlnLeuLeuProGlyIleArgGlyIleSerGlu	155				

```

; OTHER INFORMATION: Human S-adenosylhomocysteine hydrolase protein
;
; OTHER INFORMATION: sequence
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: M61831/GenBank
US-09-347-878-1

```

Alignment Scores:

Fragmented Source:					
Pred. No.:	4.2e-106	Length:	432		
Score:	1955.00	Matches:	265		
Percent Similarity:	66.33%	Conservative:	52		
Best Local Similarity:	55.4%	Mismatches:	105		
Query Match:	45.85%	Indels:	56		
DB:	4	Gaps:	5		

US-09-759-990A-1 (1-1461) X US-09-347-878-1 (1-432)

QY	37	TACAGAAATGCCGACATCAAACTCCATCTTCGGCGCCTAAGGAACCTTACCCTTGCTGAG	96
Dd		:::     :::     :::     :::     :::     :::     :::     :::	
	7	TyrsYvsAlaAspIleGlyLeuAlaAlatrPrGlyArgIysAlaLeuAspIleAlaGlu	26
QY	97	AAGGAAATGCCAGGTCTTAAGTCTTCCTGCAGCGCTATTCCGCTTCAAAGCCATCAAG	156
Dd		:::     :::     :::     :::     :::     :::     :::	
	27	AsnGluMetProGlyLeuMetArgMetArgGluArgfyrSerAlaSerLysProLeuIys	46
QY	157	GGTGTCCAGAATCTCTGGTTCCCTCCCATCAGATCCAGATCCAGACGCGCTCTCATCGACA	216
Dd		:::     :::     :::     :::     :::     :::     :::	
	47	GlyAlaArgIleAlaGlyCysLeuHisMetThrValGluThrAlaValLeuIleGluThr	66
QY	217	CTCACAGCTCTTGCTGCTGATGTCAAGTCAAGTGGCTTCTCTGCACATCTTCTTACACAGAT	276
Dd		:::     :::     :::     :::     :::     :::     :::	
	67	LeuValThrLeuGlyAlaGluValGlnTrpSerCysAsnIlePheSerThrGlnAsn	86
QY	277	ACAGCGGTGCTGCTATCGTTCTGGCCCCAACAGGCACACCAGAGAAGCCAGCGGTATC	336
Dd		:::     :::     :::     :::     :::     :::     :::	
	87	HisAlaAlaAlaAlaIleIle-----LysAlaGlyIle	97
QY	337	CAGTCTTCGCTGGAGGGCGAAACACTCCAGATACTCGGAGAACACATACCGGCT	396
Dd		:::     :::     :::     :::     :::     :::     :::	
	98	ProValTyrlAlaTrpIysGlyGluThrAspGluGluTyrLeuTrpCysIleGluGlnThr	117
QY	397	CTCACATGCCAGATGGTCAAGGCCACACAGCGTTCGATGATGGTGGTGTGCTAC	456
Dd		:::     :::     :::     :::     :::     :::     :::	
	118	LeuTyrPheLysAspGlyPro---LeuAsnMetIleLeuAspAspGlyGlyAspLeuThr	136
QY	457	CTCCTCATCTCCAAGGGCTCGAATTCGAAACAGCGGTGCTGTCCAGAGCAACAGAA	516
Dd		:::     :::     :::     :::     :::     :::     :::	
	137	AsnLeuIle-----	139
QY	517	GCTGACACCTCGAATACCGCTGCGTCTTGCTTGTCTACACTCAGCAGGTCTTCAACCAAGAC	576
Dd		:::     :::     :::     :::     :::     :::     :::	
	139	-----	139
QY	577	AAGAACCACGTGGCACACA-----GTTGCTGGCGGCATGAACGGTGTTCGGAA	624
Dd		:::     :::     :::     :::     :::     :::     :::	
	140	-----HisthrIysTyrProGlnLeuLeuProGlyIleargGlyIleSerGlu	155
QY	625	GAGACAAACAGGTGTCCACGGCTCTACACGCTCGAAGAGAGGGCAAACTCCTCTTC	684
Dd		:::     :::     :::     :::     :::     :::     :::	
	156	GluThrThrThrGlyValHisAsnLeuTyrLysMetMetalAsnGlyIleLeuIysVal	175
QY	685	CCAGCCATCAACGTCACAGACGCTGTACAAAGTCCAGATTCGATAACATCTACGGGTGC	744
Dd		:::     :::     :::     :::     :::     :::     :::	
	176	ProAlaIleasnValasnAspSerValThrylSerLysPheaspasnuLeuTyrGlyCys	195
QY	745	CGCCATCCCTTATCGATGGATCAACCGTCTCTCCGATGTCTATGATCGCGCGCAAGACA	804
Dd		:::     :::     :::     :::     :::     :::     :::	
	196	ArgGluSerLeuIleAspGlyIleIysArgAlaThrAspValMetIlealagIlysVal	215
QY	805	GCTCTCGTCAATGGGTACAGCGCATGTGGCAAGGCTGCGCTCAATCCCTCCGTGGCCAA	864
Dd		:::     :::     :::     :::     :::     :::     :::	
	216	AlaValValAlaGlyTyrGlyAspValGlyIysGlyCysAlaGlnAlaLeuargGlyPhe	235
QY	865	GGGCGTCCGGTTCATCATCACAAGAACTCGACCCAACTCGCGCTCTCCAGGTCGCCATGAA	924

[illegible]

## RESULT 5

US-08-896-005-3  
Sequence 3, Application US/08896005  
Patent No. 5854023  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Corley, Neil C.  
APPLICANT: Lal, Preeti  
APPLICANT: Shah, Purvi  
TITLE OF INVENTION: S-ADENOSYL-L-HOMOCYSTEINE HYDROLASE  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/896,005  
FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0337 US  
TELECOMMUNICATION INFORMATION:



QY	34	GAGTACAGAAATGCGCATCAACCTCCATGTTCTCGCGCGTAAGAACTTACCTTGCT	93
Db	13	AspTyrLysValAlaAspIleSerLeuAlaAspTyrGlyArgLysGluIleLysLeuAla	32
QY	94	GAGAGGAATGCCAGGCTGTTNGTCTTCTGAGCGGTATTCCGCTCTTCAAGCCATTG	153
Db	33	GluAlaGluMetProAlaLeuIleGlyLeuArgLysArgTyrAlaAlaSerLysProLeu	52
QY	154	AAGGGTGCAGAAATCTCGGTCCCTCCATGACAGTCAGACAGCCGCTCTCATCGAG	213
Db	53	AlaGlyAlaLysIleLeuGlyCysIleHisMetThrIleGlnThrAlaValLeuIleGlu	72
QY	214	ACACTCACAGCTCTGGTGGTGAATGCAGATGGGCTTCGCAACATCTTCTACACAA	273
Db	73	ThrLeuValGluLeuGlyAlaGluValArgTyrThrSerCysAsnIlePheSerThrGln	92
QY	274	GATACAGCCGCTGCTGCTTATGTTGTGCGGCCAACAGGCACACGAGAGGACGACCG	333
Db	93	AspHisAlaAlaAlaIleAlaAla-----Argly	103
QY	334	ATCCCGCTGTCGCTCGAGGGCGAAACACTCCAGAACTATCGGAGAACACATACCGC	393
Db	104	IleProValPheAlaTyrPlyGlyGlnThrGluGluLys-----	117
QY	394	GCTCTCACATGG-----CCAGATGGTCAA-----GGCCACAG	426
Db	118	-----ValTyrCysLeuGluGlnGlnIleAsnValAsnGlyGlnProTyrAspAlaAsn	135
QY	427	CAGGTGTGCATGGTGGTGATGTACACTCTCTCATCTCCAGAGGCTTCGAATCGAA	486
Db	136	MetIleLeuAspGlyGlyAspLeuThrAlaLeuValHisGluLysTyrPro-----	153
QY	487	ACAGCCGCTGTTCTCCAGACCAACAGAGAGTGACAACTCGAATACCGCTGCTGTTCT	546
Db	153	-----	153
QY	547	GCTACACTCAAGCAGCTTCTCAACCAAGACAGACCACTGGGCACACAGTTCCTCCGCG	606
Db	154	AlaLeuLeuGluArgIle-----	159
QY	607	ATGAACCGTGTTCGGAAGACAGAACAAACAGGTGTCCACCGCTCTACACAGCTCGAGAAG	666
Db	160	--HisGlyIleThrGluGluThrThrGlyValGlnArgLeuIleGluMetTrpLys	178
QY	667	CAGGCGAACTCTCTTCCAGCCATCAAGCTCACAGCGCTGTACAAAGTCCAGTTC	726
Db	179	AspGlyThrLeuLysValProAlaIleAsnValAsnAspSerValThrLysSerLysAsn	198
QY	727	GATAAATCATACGGTGGCGGCATCCCTTCATCGATGCTATCAACCGTGTCTCCGATGTC	786
Db	199	AspAsnLysTyrGlyCysArgHisSerLeuAsnAspAlaIleLysArgAlaThrAspMet	218
QY	787	ATGATCGGCGGCAGACAGCTCTCGTCATGGGTACGGCGATGTGCGCAAGGGCTCGCCT	846
Db	219	LeuLeuSerGlyArgArgAlaLeuValIleGlyTyrGlyAspValGlyLysGlySerAla	238
QY	847	CAATPCCCTCGTGGCGCAGGGCGCTCGCGTTATCATCACAGAACTCCAGCAATCTCGCCT	906
Db	239	GlnSerLeuArgGlnGluGlyMetIleValArgValThrGluValAspProIleCysAla	258
QY	907	CTCAGGCTCCCATGAGAGGCTCACAGTCC-----	936
Db	259	MetGlnAlaCysMetAspGlyTyrGluValValSerProTyrLysAsnGlyValGlnThr	278
QY	937	CGCCGCTTCAGGAAGTCTGTCAGAGATGTC-----GATATCTTCGTATACATGC	984
Db	279	GlyLysLysGluAspIleAsnHisAspLeuLeuGlyAsnThrAspLeuValThrThr	298
QY	984	ACAGAAACTCGCATATCATCTCTGTGACATGATGCGCCACGATGAAGGTAAGCTATT	1044
Db	299	ThrGlnAsnTyrHisValCysAspAlaAlaMetLeuAspSerLeuValAlaGlyValAla	318

1045	Qy	GTCCGGTACATCGGCCACTTCGATTAACGAAATGTGATACAGATGGCCTCATGAAATACCCA	1104
319	Db	ValCyeAsnIleGlyHisPheAspThrGluIleAspThrAlaTyLeuArg-----	335
1105	Qy	GGCATCAAGCACATCCCAATCAAGCCAGAA--TAGCACATGTGGAAATTCACAGAT--	1158
336	Db	GlyTyrIlyStrpValGluValIlyProGlnValHisGlnValTyArgSerGluAspGlu	355
1159	Qy	GGCCAGCGTATCCTCTTCTTGTCGAGGGCCGCTCTTTAACTTCGTGCGCTACAGGT	1218
		:::	
356	Db	AsnAsnTyrIleIleLeuLeuSerGluGlyArgLeuValAsnLeuGlyAsnAlaThrGly	375
1219	Qy	CACCCATCTTCTGTTATGTCATGTCATTTATTCACAAACAGACACACTCGCTCAGCTCGACCTC	1278
376	Db	HisProSerArgValMetAspGlySerPheAlaAsnGlnValLeuGlyGlnIleHisLeu	395
1279	Qy	TAC---GAAAGAGAGGAAATCTC-----GAGAA-----AAGGTTTAC	1314
		:::	
396	Db	PheGlnGluLysPheAlaAspLeuProAlaSerGluLysAlaAlaLysIleArgValGlu	415
1315	Qy	ACACTTCGAGACATCTCGATGAAGAAGTCGTGCGCTCCACCTCGCATCTCTCGATGTC	1374
416	Db	ValLeuProLysLeuAspGluGluValAlaAlaMetValAlaGlyPheGlyGly	435
1375	Qy	CACCTTACAAAGCTTACACAGACAGGCTGACTACATCAACGTTCCAGGTCAGGTCCT	1434
436	Db	ValLeuThrGlnLeuThrGlnGluGlnAlaAspTyrLeuGlyIleAlaValGluGlyPro	455
1435	Qy	TACAAGTCTGATGCTTACCGTTAT	1458
		:::	
456	Db	PheLysSerAspAlaTyrLysTyr	463

## RESULT 7

```

RESOLUT 7
US-09-252-991A-23501
? Sequence 23501, Application US/09252991A
? Patent No. 6551795
? GENERAL INFORMATION:
? APPLICANT: Marc J. Rubenfield et al.
? TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
? TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
? FILE REFERENCE: 107196.136
? CURRENT APPLICATION NUMBER: US/09/252,991A
? CURRENT FILING DATE: 1999-02-18
? PRIOR APPLICATION NUMBER: US 60/074,788
? PRIOR FILING DATE: 1998-02-18
? PRIOR APPLICATION NUMBER: US 60/094,190
? PRIOR FILING DATE: 1998-07-27
? NUMBER OF SEQ ID NOS: 33142
? SEQ ID NO 23501
? LENGTH: 502
? TYPE: PRT
? ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23501

```

[illegible]

383 HisArgThrGlyIysAspGlyPheAspAlaHisAsnAspAspTyrIeuIleLeuLeuAla 402  
 1183 GAGGCGCGCTCTTTAAACCTGGCTGGCTACAGGTCCACCATCTTTCGTATGTCAATG 1242  
 403 GluGlyArgLeuValAsnLeuGlyAsnAlaThrGlyHisProSerArgIleMetAspGly 422  
 1243 TCATTCAAAACACACACTCGCTCAGCTCGACCTCTACGAAAGAGA----- 1290  
 423 SerPheAlaAsnGlnValLeuAlaGlnIleHisLeuPheGluGlnIlystYrAlaAspLeu 442  
 1291 GGAAATCTCGAAG-----AAGCTTTACACACTCCGAGCATCTCGATCGAA 1338  
 443 ProAlaAlaGluIysAlaIysArgLeuSerValGluValLeuProIysIysLeuAspGlu 462  
 1339 GAAGTCGCTCGCTCCACTCGGATCTCTCGATGTCCACCTTACAAAGCTTACAGAG 1398  
 463 GluValAlaLeuGluMetValIysGlyPheGlyValValThrGlnLeuThrProIys 482  
 1399 CAGGCTGACTACATCAAGTTCACAGTTGAGGCTCTTCAAAGTCTGATGTTACCGTTAT 1458  
 483 GlnAlaGluTyrIleGlyValSerValGluGlyProPheIysProAspThrTyrArgTyr 502  
 RESULT 8  
 US-08-896-005-1  
 : Sequence 1, Application US/08896005  
 : Patent No. 5854023  
 : GENERAL INFORMATION:  
 : APPLICANT: Hillman, Jennifer L.  
 : APPLICANT: Corley, Neil C.  
 : APPLICANT: Lal, Preeti  
 : APPLICANT: Shah, Purvi  
 : TITLE OF INVENTION: S-ADENOSYL-L-HOMOCYSTEINE HYDROLASE  
 : NUMBER OF SEQUENCES: 5  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 : STREET: 3174 Porter Drive  
 : CITY: Palo Alto  
 : STATE: CA  
 : COUNTRY: USA  
 : ZIP: 94304  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Diskette  
 : COMPUTER: IBM Compatible  
 : OPERATING SYSTEM: DOS  
 : SOFTWARE: FastSeq for Windows Version 2.0  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/896,005  
 : FILING DATE: Filed Herewith  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER:  
 : FILING DATE:  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Billings, Lucy J.  
 : REGISTRATION NUMBER: 36,749  
 : REFERENCE/DOCKET NUMBER: PF-0337 US  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 415-855-0555  
 : TELEFAX: 415-845-4166  
 : INFORMATION FOR SEQ ID NO: 1:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 500 amino acids  
 : TYPE: amino acid  
 : STRANDEDNESS: single  
 : TOPOLOGY: linear  
 : IMMEDIATE SOURCE:  
 : LIBRARY: BLADTUT04  
 : CLONE: 1519044  
 US-08-896-005-1  
 Alignment Scores:  
 Fred. No.: 8,72e-86 Length: 500  
 Score: 1033.00 Matches: 210  
 Percent Similarity: 61.95% Conservative: 83

Best Local Similarity: 44.40% Mismatches: 130  
 Query Match: 37.74% Indels: 50  
 DB: 2 Gaps: 7

US-09-759-990A-1 (1-1461) x US-08-896-005-1 (1-500)

QY	43	ATTCGCCGACATCAACCTCCATGTTCTGGCGGTAAAGAACTTACCTTGTCTGAAGAA	102
		::: ::::	
Db	77	Vallylsasnllelysglnlaqlupheglvargargluileuilaalagluinasp	96
QY	103	ATGCACGCTTATGCTTCTCGTACGGCTTATTCGGCTTCTAAAGCATTCGAAGGCTGC	162
		::::	
Db	97	MetSerAlaLeuileSerLeuarglysargalacnlglyluyseProLeuAlaGlyAla	116
QY	163	AGATCTCTGGTTCCTCCACATGATGACGATCGACAGCGCTCTCATCGAGACACTCA	222
		::::	
Db	117	LysileValGlyCysThrHisleThraladlnThAlaValleuileGluThrLeuCy	136
QY	223	GCTCTTGCTGTCATGTCAGATGGGCTTCCTGCAACATCTCTCTACACAAGATACAGC	282
		::::	
Db	137	AlaIeuGlyAlaGlnCysGtrpSerAlaCysnlleryrSerThrGlnAsnGluVal	156
QY	283	GCTGCTGCTATTCGTTTCGCGCCACAGGACACACCAGAGAGCCACCGCGGTATCC	342
		::::	
Db	157	AlaAlaAlaLeuala-----GluAlaGlyValAlaVal	167
QY	343	TTCCGCTGGAAGGCGAAACACTCCAGATATCTGGGAAACACATACCGCGCTCTCA	402
		::::	
Db	168	PhelaItrpyslyGluSerGlnaspPheTrpCysIleAspArgCysValAsn	187
QY	403	TGGCCAGATCGTCAAGGCCACAGCAGGTGTTCGATGATGGTGTGATGCTACACTCC	462
		::::	
Db	188	Met---AspGlyTrpGlnAlaAsnMetIleLeuAspAspGlyGlyAspLeuThrHis	206
QY	463	ATCTCCAGGGCTTCGAATTCGAACACGCGGTGCTTCCAGAGCCCAACAGAGCTGAC	522
		::::	
Db	207	ValCyslyslustyr-----	211
QY	523	AACCTCGAATACCGCTCGCTTCTTGCTACTCAAGCAGTCTTCAACCAAGACAAGAC	582
		::::	
Db	212	-----ProAsnValPheLysLys-----	217
QY	583	CACTGGCACACAGTTGCTCGCGCANGAACGGTGTTCGAAAGACAAACACAGTGT	642
		::: ::::	
Db	218	-----IleArgGlyIleValgluIuserValThrGlyVal	229
QY	643	CACGCTCTACAGCTCGAGAAGGAGGCAAACTCTCTCCAGCGCTCAACAGTCAAC	702
		::::	
Db	230	HisargLeuTrpGlnIeuSerlysalaglylsuLeuCyValProAlaMetAsnValAs	249
QY	703	GACGCTGTTCAAAGTCCAAAGTTGATAACATCAAGGCTCGCGCCACTCCCTTATCG	762
		::::	
Db	250	AspSerValThrLysGlnlysePheAspAsnLeuTrpCysCysArgGluSerIleLeu	269
QY	763	GGTATCAACCGTCTTCGATCTCATGATCGGGGCAACAGACAGTCTCTCGTCATGGGT	822
		::::	
Db	270	GlyIeuIysargThrThrAspValMetPheGlyGlyLysGlnValValCysGlyTr	289
QY	823	GCGCATCTCGCAGGGCTCGCCTCAATCCTCCGTGGCGCAAGCGGTTCGCGTATCAT	882
		::::	
Db	290	GlyGluValGlyLysGlyCysAlaAlaLeuLysAlaLeuGlyAlaIleValTyrlle	309
QY	883	ACAGACTCGACCAATCTCGCTCTCCAGCTGCGATGGAAGGCTACAGGTCCGCGC	942
		::::	
Db	310	ThrGluIleAspProIleCysAlaLeuGlnAlaCysMetAspGlyPheArgValVal	329
QY	943	ATCGAGGAAGTCTGTCAGGATCTCGATATCTTGTTACATGTCACAGGAATCTCG	1002
		::::	
Db	330	LeuAsnGlnValIleargGlnValAspValValIleThrCysThrGlyAsnLysAsnVal	349
QY	1003	ATCTCTGTGACATGATGCGCCACAGATGAAGATAAGGCTATTGTTCGTTAACTCG	1062
		::::	
Db	350	ValThrArgGluHisAspArgMetIysAsnSerCysIleValCysAsnMetGlyHis	369

Qy	1062	TTCGATAAACGAATAATGCATACAGATTGGCCTCATCAAAATACCAGCATCACAGCACATCCCA	11232
Db	370	SerAsnThrGluLeaspValThrSerLeu---ArgThrProGluLeuThrTrpGluArg	388
Qy	1123	ATCAGCGAGAATACGACATGTGGGAATCCAGATGCCACGCTATCTCTTCTTGCT	1182
Db	389	ValArgSerGlnValAspHisValIleThrProAspGlyLysArgValValLeuLeuala	408
Qy	1183	GAGGCGCGCTCTTTAAACTTGGCTGGCTACAGGTCACCCATCTTTTCGTTATGCAATG	1242
Db	409	GluGlyArgLeuLeuAsnLeuSerCysSerThr---ValProThrPheValLeuSerIle	427
Qy	1243	TCATTACAAACGACACTCGCTTCAGCTCGACCTCTACGAA--AAGAGAGGAATCTC	1299
Db	428	ThrAlaThrThrGlnAlaLeualaLeuIleGluLeuTyraAsnAlaProGluGlyArgTyr	447
Qy	1300	GAGAAGAAGGTTTTACACACTTCCGAGCATCTCGATGAAGAAGTCGCTCGCTCCACCTC	1359
Db	448	LysGlnAspValTyrIleuLeuProLysMetAspGluTyrValAlaSerLeuHisIleu	467
Qy	1360	GGATCTCTCGATGCCACTTACAAAGCTTACACAGACGAGGCTGACTACATCAACGTT	1419
Db	468	ProSerPheAspAlaHisIleuThrGluLeuThrAspAspGlnAlaLysTyrLeuGlyLeu	487
Qy	1420	CCAGTTGAGGTCCTTCAAGCTCGATCGCTTAACCGTTAT	1458
Db	488	AsnLysasnGlyProPheLysProAsnTyrTyrArgTyr	500
 RESULT 9 US-09-252-991A-23297 ; Sequence 23297, Application US/09252991A ; Patent No. 6551795 ; GENERAL INFORMATION: ; APPLICANT: Marc J. Rubenfield et al. ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUD ; FILE REFERENCE: 107196.136 ; CURRENT APPLICATION NUMBER: US/09/252,991A ; CURRENT FILING DATE: 1999-02-18 ; PRIOR APPLICATION NUMBER: US 60/074,788 ; PRIOR FILING DATE: 1998-02-18 ; PRIOR APPLICATION NUMBER: US 60/094,190 ; PRIOR FILING DATE: 1998-07-27 ; NUMBER OF SEQ ID NOS: 33142 ; SEQ ID NO 23297 ; LENGTH: 215 ; TYPE: PRN ; ORGANISM: Pseudomonas aeruginosa US-09-252-991A-23297			
 Alignment Scores: Pred. No.: 3,89e-30 Length: 215 Score: 421.50 Matches: 114 Percent Similarity: 50.95% Conservative: 20 Best local Similarity: 43.35% Mismatches: 78 Query Match: 16.10% Indels: 51 DB: 4 Gaps: 5			
 US-09-759-990A-1 (1-1461) x US-09-252-991A-23297 (1-215)			
Qy	818	CCCATGACGAGAGCTGTCTTCGCCCGCATCATGACATCGAAGCAGCGTTGATACCAFCG	759
Db	1	ProfileThrArgAlaCysLeuProAspSerArgTyrSerValProAlaLeuMetAlaSer	20
Qy	758	ATAGGGAGGTGGCGACCCGCTAGATGTTATCGAACTTGGACTTTGTACACGGCGCTG	699
Db	21	LeuArgLeuThrParGlnProTyrIleuLeuSerPheLeuLeuValThrGlnSerLeu	40
Qy	698	ACGTTGATGGCTGGGAAGAGAGTTTGCCCTCTCTCTCGAGCTGGTAGAGCGGTGGACA	639
Db	41	ThrIleuAlaGlyThr-PheArgAlaProPheLeuSerMetSerSerArgArgTrpThr	60



```
Db 351 ----- 351
Qy 483 GAATTCGAGCCCTGGAGATGAGAGTGTAGCATCACCACCATCATCGACACCTGCTG 424
Db 352 ----- 352
Qy 423 TGGGCGCTTGACCATCGCCATGTGAGAGCGGATGTCTTCCCA---GTATCTGG 367
Db 361 GlyValProArgLeuAlaValLeuGluAspGlyLeuLeuAspAlaProPheValLeuPhe 380
Qy 366 GAGTGTTCCTCCAGCGGAGAGCTGGATACCGGCTGGCTTCTGCTGTCCTGT 307
Db 381 LeuGlyLeuAlaLeuProGlyGluHisArgAspAlaGly----- 393
Qy 306 TGGGCGCACACATAGCAGCGGCTGTATCTTGTAGAGAAGATGTTCAGGAAGC 247
Db 394 -----GlyGlyAspGlyGlyGlyLeuValLeuGlyGlyGluAspValAlaAlaArg 411
Qy 246 CMTCTGATCAGCACCAGAGCTGTGAGTGTCTCGATGAGAGCGGCTGTCTGACTGT 187
Db 412 ProAlaAspPheGlyAlaGlnArgAspGlnGlyLeuAspGlnHisAlaGlyLeuAla 431
Qy 186 CAGTGGAGGAGACAGAGATTCTGACACCTTCATGGCTTAGAGCGGAATAAGCTC 127
Db 432 HisValAspAlaThrGluAspLeuArgAlaPheGlnArgLeuLeuAlaGlyIleLeuAla 451
Qy 126 ACGAAGAACCAATAAGACTGGCATTTCTCTCAGCAAGGTAAGTTCTTACGGCCGAG 67
Db 452 AlaGlnAlaHisGlnCysGlyHisPheArgPheGlyAspAspGlnLeuAlaAlaPro 471
Qy 66 AACATGGAGGTGTATGTCGCAATTCTGTACTCGATGAGCACCAGTAGTGATTGCA 7
Db 472 GlyGlyGlnGlyAspValGlyAspLeuValValGlyGlnThrGlyArgHisAspSer 491
Qy 6 AGCCAT 1
Db 492 AlaHis 493

RESULT 11
US-08-930-894-4
; Sequence 4, Application US/08930894
; Patent No. 6037524
; GENERAL INFORMATION:
; APPLICANT: GREENLAND, Andrew James
; APPLICANT: DRAPER, John
; APPLICANT: SKIPSEY, Marc
; APPLICANT: WARNER, Simon
; TITLE OF INVENTION: S-ADENOSYL-L-HOMOCYSTEIN HYDROLASE PROMOTER
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/930,894
; FILING DATE: 09-OCT-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/00882
; FILING DATE: 10-APR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9507381.3
; FILING DATE: 10-APR-1995
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
```

```
;
; LENGTH: 140 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: ARABIDOPSIS SHH PROTEIN
US-08-930-894-4

Alignment Scores:
Pred. No.: 8,35e-24 Length: 140
Score: 351.00 Matches: 76
Percent Similarity: 62.16% Conservative: 16
Best Local Similarity: 51.35% Mismatches: 42
Query Match: 12.82% Indels: 14
DB: 3 Gaps: 3

US-09-759-990A-1 (1-1461) x US-08-930-894-4 (1-140)
Qy 280 GCCCTCTCTCTATCTGTTCGCGCCCAACAGGCACACAGAGAACCCGCGGTATCCCA 339
Db 2 AlaAlaAlaAlaIleAlaArgAspSerAlaAla----- 12
Qy 340 CTCTTCCTCTGGAAGGCGCAACACTCCAGATACTGGAGACACATACCGCGCTCTC 399
Db 13 ValPheAlaTrpGlyGluThrLeuGlnGluTrpTrpCysThrGluArgAlaLeu 32
Qy 400 ACATGGCCAGATGTCAGAGCCCGCCAGCAGCGTGTTCGATGATGTGTGTGTACACTC 459
Db 33 AspTrpGlyProGlyGlyProAspLeuIleValAspAspGlyGlyAspAlaThrLeu 52
Qy 460 CTC---ATCTCCAGGGCTTCGAA-----TTCGAAACAGCGGCTGTGTTCCTCA 504
Db 53 PheArgIleHisGlnGlyValIleAlaGluGluIlePheGluLysThrGlyGlnValPro 72
Qy 505 GAGCGCAACAGAGCTGACAACTCGAATACCGCTGCTTCTTCTTACACTCAAGCAGCTC 564
Db 73 AspProThrSerThrAspAsnProGluPheGlnIleValLeuSerIleIleLysGluGly 92
Qy 565 TTCACCAACAGCAAGAACCCACTGGCAGACAGTGTTCGCGGATGACACGGTGTTCGAA 624
Db 93 LeuGlnValAspProArgLysTrpHisLysMetLysGluArgLeuValGlyValSerGlu 112
Qy 625 GAGAACACACAGTGTCCACGCTCTACAGCTCGAGAGGAGGCGCAACTCTCTTC 684
Db 113 GluThrThrThrGlyValLysArgLeuTrpGlnMetGlnGluAsnGlyThrLeuLeuPhe 132
Qy 685 CCAGCCATCAACCTCAACGAGCT 708
Db 133 ProAlaIleAsnValAsnAspSer 140

RESULT 12
US-08-930-894-6
; Sequence 6, Application US/08930894
; Patent No. 6037524
; GENERAL INFORMATION:
; APPLICANT: GREENLAND, Andrew James
; APPLICANT: DRAPER, John
; APPLICANT: SKIPSEY, Marc
; APPLICANT: WARNER, Simon
; TITLE OF INVENTION: S-ADENOSYL-L-HOMOCYSTEIN HYDROLASE PROMOTER
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```





DB

```

Db      42  AsnGluMetProGlyLeuMetArgMetArgGluArgTyrSerAlaSerLysProLeuLys 61
Qy      157  GGHTCAGAAATCTCTGGTTCCTCCACATGACAGTCCAGACAGCGCTCCTCATCGAGACA 216
Db      62  GlyAlaArgIleAlaGlyCysLeuHisMetThrValGluThrAlaValleuIleGluThr 81
Qy      217  CTCACAGCTCTTGGTGTGATGTCAGATGGGCTTCCTGCAACATCTTC 264
Db      82  LeuValThrLeuGlyAlaGluValGlnTrpSerSerCysAsnIlePhe 97

```

Search completed: October 15, 2003, 14:56:30  
 Job time : 51 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: October 15, 2003, 14:53:35 ; Search time 97.5 Seconds  
(without alignments)  
4828.895 Million cell updates/sec

Title: US-09-759-990A-1  
Perfect score: 2737  
Sequence: 1 atggtttgcaatcacctac.....ctgatgttaccgtttattaa 1461

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 600653 seqs, 161128416 residues

Total number of hits satisfying chosen parameters: 1201306

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool/US0975990/runat 15102003 115157 2538/app query.fasta 1.1607  
-DB=Published Applications AA -OFFTX=fastan -SUFFIX=rapb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNIT5=bits -START=1 -END=1 -MATRIX=blom62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pco -NORM=ext -HEARSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USBR=US0975990 @CGN 1 1 95 @runat 15102003 115157 2538  
-NCPUS=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FCAPOP=6 -FCAPEXT=7 -YCAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	ID	Description
-----					

## ALIGNMENTS

RESULT 1  
US-10-156-761-12587  
; Sequence 12587, Application US/10156761  
; Publication NO. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OKURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 12587  
; LENGTH: 485  
; TYPE: PRT

1	1640	59.9	485	15	US-10-156-761-12587	Sequence 12587, A
2	1580	57.7	469	12	US-10-314-657-71	Sequence 71, Appl
3	1454	53.1	474	10	US-09-738-626-4336	Sequence 4336, Ap
4	1454	53.1	478	11	US-09-746-660A-98	Sequence 98, Appl
5	1454	53.1	498	11	US-09-919-854-2	Sequence 2, Appl
6	1318	48.2	432	11	US-09-746-660A-102	Sequence 102, App
7	1260	46.0	457	9	US-09-925-301-1059	Sequence 1059, Ap
8	1032	37.7	614	9	US-09-782-051-2	Sequence 2, Appl
9	958	35.0	464	12	US-10-342-224-52	Sequence 52, Appl
10	551.5	20.1	170	12	US-10-040-895-14	Sequence 14, Appl
11	541	20.7	236	10	US-09-738-626-4337	Sequence 4337, Ap
12	426.5	16.3	240	10	US-09-738-626-4338	Sequence 4338, Ap
13	287	10.5	97	9	US-09-799-946-9	Sequence 9, Appl
14	271	9.9	95	9	US-09-799-946-7	Sequence 7, Appl
15	219	8.0	2283	12	US-10-172-502-4	Sequence 4, Appl
16	213	7.8	528	12	US-09-840-746-20	Sequence 20, Appl
17	208	7.6	2344	9	US-09-815-242-12713	Sequence 12713, A
18	207	7.6	560	9	US-09-815-242-13057	Sequence 13057, A
19	205	7.5	502	9	US-09-815-242-5904	Sequence 5904, Ap
20	197	7.2	1236	11	US-09-769-787-109	Sequence 109, App
21	177.5	6.5	605	10	US-09-801-368-428	Sequence 1, Appl
22	172.5	6.3	946	12	US-09-840-746-1	Sequence 40, Appl
23	172.5	6.3	1481	10	US-09-371-900-40	Sequence 60, Appl
24	172.5	6.3	1481	10	US-09-324-417-60	Sequence 40, Appl
25	172.5	6.3	1481	12	US-10-186-950-40	Sequence 40, Appl
26	166	6.1	596	11	US-09-946-374-243	Sequence 243, App
27	166	6.1	596	12	US-10-015-387A-243	Sequence 243, App
28	166	6.1	596	12	US-10-063-735-100	Sequence 100, App
29	166	6.1	596	12	US-10-006-130A-243	Sequence 243, App
30	166	6.1	596	12	US-10-199-672-310	Sequence 310, App
31	166	6.1	596	12	US-10-006-172A-243	Sequence 243, App
32	166	6.1	596	12	US-10-187-749-310	Sequence 310, App
33	166	6.1	596	12	US-10-194-457-310	Sequence 310, App
34	166	6.1	596	12	US-10-184-642-310	Sequence 310, App
35	166	6.1	596	12	US-10-196-747-310	Sequence 310, App
36	166	6.1	596	12	US-10-015-392A-243	Sequence 243, App
37	166	6.1	596	12	US-10-017-253A-243	Sequence 243, App
38	166	6.1	596	12	US-10-173-689-310	Sequence 310, App
39	166	6.1	596	12	US-10-173-690-310	Sequence 310, App
40	166	6.1	596	12	US-10-173-691-310	Sequence 310, App
41	166	6.1	596	12	US-10-173-692-310	Sequence 310, App
42	166	6.1	596	12	US-10-173-694-310	Sequence 310, App
43	166	6.1	596	12	US-10-173-698-310	Sequence 310, App
44	166	6.1	596	12	US-10-173-699-310	Sequence 310, App
45	166	6.1	596	12	US-10-173-707-310	Sequence 310, App



Db 11 AspPheLysValAlaAspLeuSerLeuAlaAlaPheGlyArgLysGluIleThrLeuAla 30  
Qy 94 GAGAGAAATGCCAGGTCTTATGGTTCTTCGTGAGCGTTATTCGCGTTCCTAAGCCATTG 153  
Db 31 GluHisGluMetProGlyLeuMetSerIleArgGluGluIleAlaAlaGlnProLeu 50  
Qy 154 AGGGTGTGCAAGATCTCTGGTTCCTCCATGACATGTCAGACAGCGTCTCTCATCAG 213  
Db 51 AlaGlyAlaArgIleThrGlySerLeuHisMetThrValGlnThrAlaValLeuIleGlu 70  
Qy 214 ACACCTCACAGCTCTTGTGTGTATGTCAGATGGCTCTCCGCAACATCTCTCTACAA 273  
Db 71 ThrLeuValAlaLeuGlyAlaAspValArgTrpValSerCysAsnIleTrpSerThrGln 90  
Qy 274 GATACAGCGCTGCTGCTATCGTTGTCGCCCAACAGAGCACACAGAGAGAGCGCGGT 333  
Db 91 AspHisAlaAlaAlaAlaAlaAla-----AlaGly 101  
Qy 334 ATCCAGCTTCGCTCGAGGCGAAACACTCCAGAACTACCTGGAGAGACATACCGC 393  
Db 102 IleProValPheAlaTrpLysGlyGluThrLeuGluIleTrpCysThrGluGln 121  
Qy 394 GCTCTCACATGCCAGATGTCGAGGCCACACAGAGGTTCGATGATGTCGTCATGCT 453  
Db 122 AlaLeuThrTrpProGlyHisThrGlyProAsnMetIleLeuAspaspGlyGlyAspAla 141  
Qy 454 ACATCTCATCTCCAGGGTTCGAATTCGAAACAGCGGGTGTGTCAGAGGCAACA 513  
Db 142 ThrLeuLeuValHisLysGlyValGluTrpArgLysThrGlyIleLeuPro----- 158  
Qy 514 GAAGTGCACACCTCGAATACCGCTGCTGTCGATGTCGTCACCTCAAGCAGGTCTTCAACAA 573  
Db 159 GluAlaGluAsnGluGlu-----LeuAlaValAlaArgAlaLeuLeuAspArg 174  
Qy 574 GACAAAGAACCTCGACACAGTGTGTCGCCGATGACAGCGTGTTCGAGAGAGACAACA 633  
Db 175 SerGlyLeuAspTrpThrAlaMetSerSerGluIleArgGlyValThrGluGluThrThr 194  
Qy 634 ACAGGTGCCACCGCTCTACAGCTCGAGAGAGGCGCAACTCTCTCTCCAGCCATC 693  
Db 195 ThrGlyValHisArgLeuTrpGluMetHisArgaspGlyThrLeuLeuPheProAlaIle 214  
Qy 694 AACGTCAACAGCGTGTTCACAAAGTCCAAAGTTCGATACATCTACGGCTCGCGCCACTCC 753  
Db 215 AsnValAsnAspAlaValThrLysSerLysPheAspAsnLysTrpGlyCysArgHisSer 234  
Qy 754 CTTATCGATGTATCAACCGTGTCTCCGATGTCATGTCGCGGCAAGACAGCTCTCGTC 813  
Db 235 LeuIleAspGlyIleAsnArgAlaThrAspValLeuIleGlyLysThrAlaValVal 254  
Qy 814 ATGGGTACGGGTGCGGCAAGGCTGCGCTCATCTCCGTCGCGCAAGCGCTCGC 873  
Db 255 CysGlyTrpGlyAspValGlyLysGlyCysAlaGluSerLeuArgGlyGlnGlyAlaArg 274  
Qy 874 GTTATCATCAGAACTCGACCAATCTCGCTCTCCAGGCTGCCATGGAAGCTTACCAG 933  
Db 275 ValIleIleThrGluIleAspProIleCysAlaLeuGlnAlaAlaMetaspGlyTrpGln 294  
Qy 934 GTCCGCGCAGTCGAGAGTCTGTCAGAGGATGTCGATCTTGGTTTCATGTCACAGAAAC 993  
Db 295 ValAlaThrLeuAspGluValValAspLysAlaAspIlePheValThrThrThrGlyAsn 314  
Qy 994 TGGGATATCATCTCTGTGATGATGATGCGCCAGATGAGGATAGGCTATCTCGGTAC 1053  
Db 315 LysAspIleIleMetAlaSerAspMetAlaLysMetLysHisGlnAlaIleValGlyAsn 334  
Qy 1054 ATCCGCCATTCGATACGAATTCATGATACAGATGGCTCATGAATACCCAGCATCAAG 1113  
Db 335 IleGlyHisPheAspAsnGluIleAspMetAlaGlyLeuAlaGlnIleProGlyIleVal 354  
Qy 1114 CATATCCCATCAGCAGAAATACGATGTGGGAATTCACAGATGCCACGCTATCCTC 1173  
Db 355 LysAspGluValLysProGlnValHisThrTrpLysPheProAspGlyLysValLeuIle 374

Qy 1174 CTTCTTGTGAGGCGCGCTTCTTAACCTTGGCTGCGCTACAGGTCAACCATCTTTGCTT 1233  
Db 375 ValLeuSerGluGlyArgLeuLeuAsnLysGlyAsnAlaThrGlyHisProSerPheVal 394  
Qy 1234 ATGTCATATGTCATTCACAAACAGACACTCGCTCAGCTGACCTCTACGAAAGAGAGGA 1293  
Db 395 MetSerAsnSerPheAlaAspGlnThrLeuAlaGlnIleGluLeuPheThrLysProAsp 414  
Qy 1294 AATCTCGAGAGAGGTTTACACACTTCCGAGGACATCTCGATGAGAAAGTCTCGCTC 1353  
Db 415 GluTrpProThrAspValTrpValLeuProLysHisLeuAspGluLysValAlaArgLeu 434  
Qy 1354 CACCTCGGATCTCTCGATCTCCACCTTACAAAGCTTACACAGAGCAGCGCTCATCATC 1413  
Db 435 HisLeuAspAlaLeuGlyValLysLeuThrThrLeuArgProGluGlnAlaSerTrpIle 454  
Qy 1414 AACGTTCCAGTTGAGGTCCTTACAGTCTGATGCTTACCGTTAT 1458  
Db 455 GlyValGluValAspGlyProTrpLysProAspHisTrpArgTrp 469

## RESULT 3

US-09-738-626-4336  
; Sequence 4336, Replication US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: Patent In ver. 3.0  
; SEQ ID NO 4336  
; LENGTH: 474  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4336

## Alignment Scores:

Pred. No.:	1 37e-109	Length:	474
Score:	1454.00	Matches:	282
Percent Similarity:	75.84%	Conservative:	79
Best Local Similarity:	59.24%	Mismatches:	111
Query Match:	53.12%	Indels:	4
DB:	10	Gaps:	4

US-09-759-990A-1 (1-1461) x US-09-738-626-4336 (1-474)

Qy 34 GAGTACAGAAATCCGACATCAACCTCATGTTCTCGCGCGTAAGAACTTACCTTTGCT 93  
Db 2 AspPheLysValAlaAspLeuSerLeuAlaGluAlaGlyArgHisGlnIleArgLeuAla 21  
Qy 94 GAGAGAAATGCGCAGGTCTTATGTTCTTCGTCGAGCGTTATTCGCTTCTAAGCATTC 153  
Db 22 GluTrpGluMetProGlyLeuMetGlnLeuArgLysPheAlaAspGluGlnProLeu 41  
Qy 154 AAGGTGTGAGATCTCTGTTCCCTCCACATGACAGTCCAGACAGCGCTCTCATCGAG 213

Db	42	LySGLYALAArgILleAlAGlySerILleHISMetThrValGlnThrAlaValLeuILleGlu	61
QY	214	ACACTCACAGCTCTTGGTGGTGTGATGTCAGATGGGCTCTCTGCACACATCTTCTACACAA	273
Db	62	ThrLeuThrAlaLeuGlyAlAGluValArgTrpAlaSerCysAsnIlePheSerThrGln	81
QY	274	GATACAGCCGCTGCTGCTATCGTTGTGCGCCACACAGCACAACAGAGAACGACGCGT	333
Db	82	AspGluAlaAlaAlaIleValValGly--SerGlyThrValGluGluProAlaGly	100
QY	334	ATCCACAGTCTTCGCTGCGAGAGGGCGAAACACATCCACAGATACTGGGAGAACACATACCGC	393
Db	101	ValProValPheAlaTrpIysGlyGluSerLeuGluGluTrpTrpCysIleAsnGln	120
QY	394	GCTCTCACATGCCAGATGGTCAAGGCCACACAGCGATTTCGATCATGGTGGTGA	453
Db	121	IlePheSerTrpGlyAsp--GluLeuProAsnMetIleLeuAspAspGlyGlyAspAla	139
QY	454	ACACTCTCTCATCTCCAGAGGCTTCGATTCGAAACACAGCGGTGCTCTTCCAGAGCCACACA	513
Db	140	ThrMetAlaValIleArgGlyA--gGluTrpGluGlnAlAGlyLeuValPro---ProAla	158
QY	514	GAAGCT---GACAACTCGAATACCGCTCGTCTTGTCTACACTCAAGCAGCTCTTCAAC	570
Db	159	GluAlaAsnAspSerAspGluTyrIleAlaPheLeuGlyMetLeuArgGluValLeuAla	178
QY	571	CAACACAGAACCACTGSCACACAGTGTGTCGGCGCATGAACGGTGTTCGAGAGACA	630
Db	179	AlaGluProGlyIysTrpGlyLysIleAlaGluAlaValLysGlyValThrGluGluThr	198
QY	631	ACAACAGGTCTCCACGGCTCTACACAGCTCGAAGAGGAGGCAAACTCCTCTTCCAGCC	690
Db	199	ThrThrGlyValHisArgLeuTyrHisPheAlaGluGluGlyValLeuProPheProAla	218
QY	691	ATCAAGCTCAACAGACGCTGTTTACAAAGTCCAAAGTTCGATTAACATCTACGGCTGCGGCAC	750
Db	219	MetAsnValAsnAspAlaValThrLysSerLysPheAspAsnLysTyrGlyThrArgHis	238
QY	751	TCCCTTATCGATGGTATCAACCGTCTCCGATGTCATGATCGCGCGCAAGACAGCTCTC	810
Db	239	SerLeuIleAspGlyIleAsnArgAlaThrAspMetLeuMetGlyGlyLysAsnValLeu	258
QY	811	GTCATGGGTTACGGCGATGTCGGCAAGGCTCGCTCAATCCCTCGCTGGCGCAAGCGCT	870
Db	259	ValCysGlyTyrGlyAspValGlyLysGlyCysAlaGluAlaPheAspGlyGlnGlyAla	278
QY	871	CGCTTATCATCAAGACTCGACCCAACTCGGCTCTCCAGGCTGCGCATGAAGGCTAC	930
Db	279	ArgValLysValThrGluAlaAspProIleAsnAlaLeuGlnAlaLeuMetAspGlyTyr	298
QY	931	CAGTCCGCGCATCGAGAAAGTGTCAAGATGTGATATCTTCGTATCATGCACAGGA	990
Db	299	SerValValThrValAspGluAlaIleGluAspAlaAspIleValIleThrAlaThrGly	318
QY	991	AACTCGATATCATCTCTGTTCATCATGTGCGCCAGATGAAGGATAAGCTTATTCGGT	1050
Db	319	AsnLysAspIleIleSerPheGluGlnMetLeuLysMetLysAspHisAlaLeuLeuGly	338
QY	1051	AACATCGGCCACTTCGATACGAATATGATACAGATGGCTCATGAATACCGAGCAT	1110
Db	339	AsnIleGlyHisPheAspAsnGluIleAspMetHisSerLeuLeuHisArgAspAspVal	358
QY	1111	AGACACATCCCAATCAAGCCAGCATACGATGTGGGAATTCGACAGATGGCCACGCTATC	1170
Db	359	ThrArgThrThrIleLysProGlnValAspGluPheThrPheSerThrGlyArgSerIle	378
QY	1171	CTCCTTCCTGTCGAGGCGCGCTCTTAACTCTGGTGGCTGCTACAGCTACACCATCTTTC	1230
Db	379	IleValLeuSerGluGlyArgLeuMetLeuGlyAsnAlaThrGlyHisProSerPhe	398
QY	1231	GTATTGTCAATGTCAATCAAAACACACACTCGTTCAGCTCGACCTTACGAAAGAGA	1290

Db 399 ValMetSerAsnSerPheAlaAspGlnThrIleAlaGlnIleGluLeuPheGlnAsnGlu 418  
 QY 1291 GGAATTCGAGAGAGAAGCTTTACACACTTCCGAAGCACTTCGATGAAAGATCGCTCGC 1350  
 Db 419 GlyGlnTyrGluAsnGluValTyrArgLeuProLysValLeuAspGluLysValAlaArg 438  
 QY 1351 CTCACCTCGGATCTCTCGATGCCATCCACCTTACAAAGCTTACACAGAGCAGGCTGACTAC 1410  
 Db 439 IleHisValGluAlaLeuGlyGlyGlnLeuThrGluLeuThrLysGluGlnAlaGluTyr 458  
 QY 1411 ATCAAGCTTCCAGTTTCAGGGTCCCTTACAAGTCTGTACGTCTTACCGTTAT 1458  
 Db 459 IleGlyValAspValAlaGlyProPheLysProGluHisTyrArgTyr 474

RESULT 4  
 US-09-746-660A-98  
 ; Sequence 98, Application US/09746660A  
 ; Publication No. US20030049804A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Pompeius, Markus  
 ; APPLICANT: Kroger, Burkhard  
 ; APPLICANT: Schroeder, Hartwig  
 ; APPLICANT: Zelder, Oskar  
 ; APPLICANT: Haberhauer, Gregor  
 ; APPLICANT: Kim, Jun-Won  
 ; APPLICANT: Lee, Heung-Schick  
 ; APPLICANT: Hwang, Byung-Joon  
 ; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING  
 ; TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS  
 ; FILE REFERENCE: BGI-121CP2  
 ; CURRENT APPLICATION NUMBER: US/09/746,660A  
 ; CURRENT FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 09/606740  
 ; PRIOR FILING DATE: 2000-06-23  
 ; PRIOR APPLICATION NUMBER: 09/603124  
 ; PRIOR FILING DATE: 2000-06-23  
 ; PRIOR APPLICATION NUMBER: 60/141031  
 ; PRIOR FILING DATE: 1999-06-25  
 ; PRIOR APPLICATION NUMBER: 60/142101  
 ; PRIOR FILING DATE: 1999-07-02  
 ; PRIOR APPLICATION NUMBER: 60/148613  
 ; PRIOR FILING DATE: 1999-08-12  
 ; PRIOR APPLICATION NUMBER: 60/187970  
 ; PRIOR FILING DATE: 2000-03-09  
 ; PRIOR APPLICATION NUMBER: DE 199311420.9  
 ; PRIOR FILING DATE: 1999-07-08  
 ; NUMBER OF SEQ ID NOS: 125  
 ; SOFTWARE: PatentIn Vers. 2.0  
 ; SEQ ID NO 98  
 ; LENGTH: 478  
 ; TYPE: PRT  
 ; ORGANISM: Corynebacterium glutamicum  
 US-09-746-660A-98

Alignment Scores:  
 Pred. No.: 1 37e-103 Length: 478  
 Score: 1454.00 Matches: 282  
 Percent Similarity: 75.84% Conservative: 79  
 Best Local Similarity: 59.24% Mismatches: 111  
 Query Match: 53.12% Indels: 4  
 DB: 11 Gaps: 4

US-09-759-990A-1 (1-1461) x US-09-746-660A-98 (1-478)  
 QY 34 GAGTACAGAAATTCGACATCAACTCCATGTCTTCGGCCGTAAGGAACCTTACCTTGCT 93  
 Db 6 AspPheLysValAlaAspLeuSerLeuAlaGluAlaGlyArgHisGlnIleArgLeuAla 25  
 QY 94 GAGAGGAATGCCAGGTCTTATGGTCTTCCTCGTACGCTTATTCGCTTCTTAAGCCATTG 153  
 Db 26 GluTyrGluMetProGlyLeuMetGlnLeuArgLysGluLysPheAlaAspGluGlnProLeu 45  
 QY 154 AAGGGTGTCAAGATCTCTGGTTCCTCCATGACAGTCCACAGACGGCTCTCATCGAG 213

46	LySGlyAlaArgIleAlaGlySerIleHisMetThrValGlnThrAlaValLeuIleGlu	65
214	ACACTCACAGCTCTTGGTGGTGAATGATGCAGATGGGCTCTCGCAACATCTTCTTACACAA	273
66	ThrLeuThrAlaLeuGlyAlaGluValArgTTPalaSerCysAsnIlePheSerThrGln	85
274	GATACAGCGCTGCTGCTATCGTTGTGCGGCCCAACAGGCACACAGAGAACCGACGCGT	333
86	AspGluAlaAlaAlaIleValValGly--SerGlyThrValGluGluProAlaGly	104
334	ATCCCACTCTTCGCTCGAAGGGCGAAACACTCCAGAACTACTGGAGAACACATACCGC	393
105	ValProValPheAlaTrpLysGlyGluSerLeuGluGluIleTrpCysIleAsnGln	124
394	GCTCTCCATGCGCCAGATGGTCTCAAGGCCACACAGAGGTTGCGATGATGGTGGTATGCT	453
125	IlePheSerTrpGlyAsp--GluLeuProAsnMetIleLeuAspAspGlyGlyAspAla	143
454	ACACTCTCATCTCCAGAGGCTTCGAATTCGAAACACAGCGGTGCTGTCCAGAGCCACAA	513
144	ThrMetAlaValIleArgGlyArgGluTrpGluGlnAlaGlyLeuValPro--ProAla	162
514	GAAGCT---GACAACTCCGAATACCGCTCGTCTTCTGTACACTCAAGCAGCTTCTCAAC	570
163	GluAlaAsnAspSerAspGluTrpIleAlaPheLeuGlyMetLeuArgGluValLeuAla	182
571	CAAGACAAAGAACCACTGCCACACAGTGTGTCGGCGCATCAACGGTGTTCGAGAGACAA	630
183	AlaGluProGlyLysTrpGlyLysIleAlaGluAlaValLysGlyValThrGluGluThr	202
631	ACAACAGGTGTCCACCGCTCTACCAAGCTCGAAGAGAGGCGCAAACTCCTCTTCCAGCC	690
203	ThrThrGlyValHisArgLeuTrpHisPheAlaGluGluGlyValLeuProPheProAla	222
691	ATCAACGTCAAGACAGCTGTTCAAAGTCAAGTTCGATTAACATCTACGGCTCGCGCAC	750
223	MetAsnValAsnAspAlaValThrLysSerLysPheAspAsnLysTrpGlyThrArgHis	242
751	TCCCTTATCATGATGGTATCAACCGTCTCCGATGTCAATGTCGCGCGGCAAGACAGCTCTC	810
243	SerLeuIleAspGlyLysAsnArgAlaThrAspMetLeuMetGlyGlyLysAsnValLeu	262
811	GTATGGGTATACGGCATGTCCGACAGGGCTCGCTCAATCCCTCGTGGCCCAAGCGCT	870
263	ValCysGlyTrpGlyAspValGlyLysGlyCysAlaGluAlaPheAspGlyGlnGlyAla	282
871	CGCTTATCATACACAGAACTCGACCAATCTCGCTCTCCAGGCTGCCATGGAAGGCTAC	930
283	ArgValLysValThrGluAlaAspProlLeuAlaLeuGlnAlaLeuMetAspGlyTyr	302
931	CAGTCTCGCGCATCGAGGAAGTCGTCAAGATGTCCGATATCTTCGTATCATGCACAGGA	990
303	SerValValThrValAspGluAlaIleGluAspAlaAspIleValIleThrAlaThrGly	322
991	AACGTCGATATCATCTCTGTGACATCATGGCCCGCAGATGAAGATTAAGCTATTGCGGT	1050
323	AsnLysAspIleIleSerPheGluGlnMetLeuLysMetLysAspHisAlaLeuLeuGly	342
1051	AACATCGGCCACTTCGATAACGAATATGATACAGATGCCCTCATGAAATACCCAGGCAT	1110
343	AsnIleGlyHisPheAspAsnGluIleAspMetHisSerLeuLeuHisArgAspAspVal	362
1111	AAGCACATCCCAATCAAGCCAGATACGACATGTGGGAATTCACAGATGGCCACCGCTATC	1170
363	ThrArgThrThrIleLysProGluValAspGluPheThrPheSerThrGlyArgSerIle	382
1171	CTCCTCTCTGCTGAGGCGCGCTTCTTAACTTGGCTGGCTACAGCTCACCCATCTTTC	1230
383	IleValLeuSerGluGlyArgLeuLeuAsnLeuGlyAsnAlaThrGlyHisProSerPhe	402
1231	GTTATGTCAATGTCAATTCACAAACACACACTCGCTCGAGCTCTTACGAAAGAGA	1290

Db 403 ValMetSerAsnSerPheAlaAspGlnThrIleAlaGlnIleGluLeuPheGlnAsnGlu 422  
 QY 1291 GGAAATCTCGAAGAAGAGGTTTACACACTTCCGAAGCATCTCGATGAAGAAGTCGCTCGC 1350  
 Db 423 GlyGlnTyAGluAspGluValTyArgLeuProLysValLeuAspGluLysValAlaArg 442  
 QY 1351 CTCACCTCGGATCTCTCGATGCCACCTTACAAGAGCTTACACAGAAGCAGCGCTGACTAC 1410  
 Db 443 IleHisValGluAlaLeuGlyGlyGlnLeuThrGluLeuThrLysGluGlnAlaGluTy 462  
 QY 1411 ATCAAGCTTCCAGTTGAGGGTCCCTTACAAGTCTGATGCTTACCGTTAT 1458  
 Db 463 IleGlyValAspValAlaGlyProPheLysProGluHisTyArgTy 478

RESULT 5  
 US-09-919-854-2  
 ; Sequence 2, Application US/09919854  
 ; Publication No. US20030100080A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FURTHWICK, MIKE  
 ; APPLICANT: HUTHMACHER, KLAUS  
 ; APPLICANT: BREHME, JENNIFER  
 ; APPLICANT: PFEFFERLE, WALTER  
 ; APPLICANT: BINDER, MICHAEL  
 ; APPLICANT: GREISSINGER, DIETER  
 ; APPLICANT: THIERSCH, GEORG  
 ; TITLE OF INVENTION: Nucleotide sequences which code for the sahH gene  
 ; FILE REFERENCE: 211739USOX  
 ; CURRENT APPLICATION NUMBER: US/09/919,854  
 ; CURRENT FILING DATE: 2002-12-30  
 ; PRIOR APPLICATION NUMBER: DE 100 44 706.6  
 ; PRIOR FILING DATE: 2000-09-09  
 ; PRIOR APPLICATION NUMBER: DE 101 09 685.2  
 ; PRIOR FILING DATE: 2001-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/294,277  
 ; PRIOR FILING DATE: 2001-05-31  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 2  
 ; LENGTH: 498  
 ; TYPE: PRT  
 ; ORGANISM: Corynebacterium glutamicum  
 US-09-919-854-2

Alignment Scores:  
 Pred. No.: 1,4e-109 Length: 498  
 Score: 1454.00 Matches: 282  
 Percent Similarity: 75.84% Conservative: 79  
 Best Local Similarity: 59.24% Mismatches: 111  
 Query Match: 53.12% Indels: 4  
 DB: 11 Gaps: 4

US-09-759-980A-1 (1-1461) x US-09-919-854-2 (1-498)

QY 34 GAGTACAGATTCCGACATCACTCCATGTTCTCGGCCGTAAAGGAACCTTACCCTGCT 93  
 Db 26 AspPheLysValAlaAspLysSerLeuAlaGluAlaGlyArgHisGlnIleArgLeuAla 45  
 QY 94 GAGAAGGAATGCCAGGTCTTATGGTTCTTCGTGACGGCTTATCCGCTTCTAAGCCATTG 153  
 Db 46 GluTyrgluMetProGlyLeuMetGlnLeuArgLysGluPheAlaAspGluGlnProLeu 65  
 QY 154 RAGGGTGTCAAGATCTCTGGTTTCCCTCCATGACAGTCCAGACAGCGCTCTCATCGAG 213  
 Db 66 LysGlyAlaArgIleAlaGlySerIleHisMetThrValGlnThrAlaValLeuIleGlu 85  
 QY 214 ACATCTCAACACTCTGGTGTGCTGATGTCAGATGGGTTCTCTCAACATCTTCTTACACAA 273  
 Db 86 ThrLeuThrAlaLeuGluValAlaGluValArgTyrAlaSerCysAsnIlePheSerThrGln 105  
 QY 274 GATACAGCCGCTGCTACTGTTGTTCGGCCCAACAGGCAACACAGAGAAGCCAGCCGGT 333  
 Db 106 AspGluAlaAlaAlaIleValAlaGlyLysSerGlyThrValGluGluProAlaGly 124







```
Db 201 ProAlaIleAsnValAsnAspSerValThrLysSerLysPheAspAsnLeuTyrGlyCys 220
QY 745 CGCCACTCCCTTATCGATGATATCAACCGTCTCCGATGTCATGATCGCGGCGAAGACA 804
Db 221 ArgGluSerLeuIleAspGlyIleLysArgAlaThrAspValMetIleAlaGlyLysVal 240
QY 805 GCTCTCGTCATGCGGTACGGCGATGCGCAAGCGCTCGCTCAATCCCTCCGTCGCCAA 864
Db 241 AlaValAlaIleGlyTyrGlyAspValGlyLysGlyCysAlaGlnAlaLeuArgGlyPhe 260
QY 865 GGGCGCTCGGTATATCATCAAGAACTCGACCAACTCGCTCGCTCCAGGTCGCCATGAA 924
Db 261 GlyAlaArgValIleIleThrGluIleAspProIleAsnAlaLeuGlnAlaAlaMetGlu 280
QY 925 GCGTACAGCTCCGCGCATCGAGGAAGTCGCAAGGATGCGATATCTTCGTTACATGC 984
Db 281 GlyTyrGluValThrThrMetAspGluAlaCysGlnGluGlyAsnIlePheValThrThr 300
QY 985 ACAGGAACCTGCGATATCATCTCTGTGTGACATGATGGCCCGAGATGAAGGATATT 1044
Db 301 ThrGlyCysIleAspIleIleLeuGlyArgHisPheGluGlnMetLysAspAlaIle 320
QY 1045 GTCGGTAACATGCGCACTTCGATACGAAATTGATACAGATGCGCTCATGAATACCCA 1104
Db 321 ValCysAsnIleGlyHisPheAspValGluIleAspValLysTrpLeuAsnGlu---Asn 339
QY 1105 GGCATCAAGCACATCCCAATCAAGCCAGAAATACGATGTGGGAATTCOCAGATGCCAC 1164
Db 340 AlaValGluLysValAsnIleLysProGlnValAspArgTyrArgLeuLysAsnGlyArg 359
QY 1165 GCTATCCTCTTCTGCTAGGCGCGCTTCTTACCTTGGCTCGCTACAGTCAACCA 1224
Db 360 ArgIleIleLeuAlaGluGlyArgLeuValAsnLeuGlyCysAlaMetGlyHisPro 379
QY 1225 TCTTTCTGTTATGCAATCTCATTCACAAACACAGACACTCGCTCAGCTCGACCTCTACGA 1284
Db 380 SerPheValMetSerAsnSerPheThrAsnGlnValMetAlaGlnIleGluLeuTrpThr 399
QY 1285 AAGAGAGGAATCTCGAAGAAGTTTACACATTCGCAAGCATCTCGATGAGAAGTC 1344
Db 400 HisProAspLysTyrProValGlyValHisPheLeuProLysLysLeuAspGluAlaVal 419
QY 1345 GCTCGCTCCATCGATCTCTCGATGTCACCTTACAAAGCTTACAGAGCAGGCT 1404
Db 420 AlaGluAlaHisLeuGlyLysLeuAsnValLysLeuThrLysLeuThrGluLysGlnAla 439
QY 1405 GACTATCAACAGTTCCAGTTGAGGTCCTTACAAAGTCTGATGCTTACCGTTAT 1458
Db 440 GlnTyrLeuGlyMetSerCysAspGlyProPheLysProAspHisTyrArgTyr 457
```

RESULT 8

```
US-09-782-051-2
; Sequence 2, Application US/09782051
; Patent No. US20020035078A1
; GENERAL INFORMATION:
; APPLICANT: Hart, Derek N J
; TITLE OF INVENTION: Enzyme having S-adenosyl-L-homocysteine hydrolase
; TITLE OF INVENTION: (AHCY) type activity
; FILE REFERENCE: 24305 MRB
; CURRENT APPLICATION NUMBER: US/09/782,051
; CURRENT FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: PCT/NZ97/00133
; PRIOR FILING DATE: 1997-10-06
; PRIOR APPLICATION NUMBER: NZ 299507
; PRIOR FILING DATE: 1996-10-04
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 614
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-782-051-2
```

Alignment Scores:

```
Pred. No.: 3,34e-75 Length: 614
Score: 1032.00 Matches: 210
Percent Similarity: 61.95% Conservative: 83
Best Local Similarity: 44.40% Mismatches: 130
Query Match: 37.71% Indels: 50
DB: 9 Gaps: 7
US-09-759-990A-1 (1-1461) x US-09-782-051-2 (1-614)
QY 43 ATTGCCGACATCAACCTCCATCTTCGCGCTAAGGAACCTTACCTTCTGAGAGGAA 102
Db 191 VallysAsnIleLysGlnAlaGluPheGlyArgArgGluIleGluIleAlaGluInAsp 210
QY 103 ATGCAGAGCTTTATGTTCTTCGTGAGCGTTATTTCGCTTCTAAGCCATTGAAGGGTGC 162
Db 211 MetSerAlaLeuIleSerLeuArgLysArgAlaGlnGlyLysProLeuAlaGlyAla 230
QY 163 AGAATCTCTGTTCCCTCCACATGACAGTCCAGACAGCGCTCTCATCGACACTCACA 222
Db 231 LysIleValGlyCysThrHisIleThrAlaGlnThrAlaValIleGluThrLeuCys 250
QY 223 GCTCTTGTGTCATGTCAGATGGCTTCTCGCAACATCTTCTCTACACAAGATACAGCC 282
Db 251 AlaLeuGlyAlaGlnCysArgTrpSerAlaCysAsnIleTyrSerThrGlnAsnGluVal 270
QY 283 GCTGCTGCTATCTGTTCTGCGCCCAACAGGCACACAGAGAACGACGCGTATCCCCAGTC 342
Db 271 AlaAlaAlaLeuAla-----GluAlaGlyValAlaVal 281
QY 343 TTCGCTGGAAGGGCAAAACATCCACAGATCTGGGAGAACACATACGGCGCTCTCACA 402
Db 282 PheAlaTrpLysGlyGluSerGluAspAspPheTrpTyrCysIleAspArgCysValAsn 301
QY 403 TGGCCAGCATGTCTAAGCCCAACAGCAGCTTTCGATGATGGTGGTGTGATCTACTCCTC 462
Db 302 Met---AspGlyTrpGlnAlaAsnMetIleLeuAspAspGlyGlyAspLeuThrHisTrp 320
QY 463 ATCTCCAAAGGGTTCGAATTCGAATTCGAACAGCGGTCTGTTCCAGAGCCACAGAGCTGAC 522
Db 321 ValTyrLysLysTyr-----ProAsnValPheLysLys----- 325
QY 523 AACCTCGAATACCGCTCGCTTCTTCTGCTACTACAGCAGGTCTTCAACCAAGACAAGAAC 582
Db 326 -----ProAsnValPheLysLys----- 331
QY 583 CACTGGCACACAGTTGCTGCGGCATGAACGGTGTTCGGAAGAGACAAACAGGTTGC 642
Db 332 -----IleArgGlyIleValGluGluSerValThrGlyVal 343
QY 643 CACCGCTCTACAGCTCCAGAGAGGAGGCAACTCTCTTCCAGCCATCAAGCTCAAC 702
Db 344 HisArgLeuTyrGlnLeuSerLysAlaGlyLysLeuCysValProAlaMetAsnValAsn 363
QY 703 GACGCTGTTCACAAAGTCCAAAGTTCGATTAACATCTACGCTGCGCCACTCCCTTATCGAT 762
Db 364 AspSerValThrLysGlnLysPheAspAsnLeuTyrCysCysArgGluSerIleLeuAsp 383
QY 763 GGTATCAACCGTCTCCGATGTCATGATCGCGCGCAAGACAGCTCTCTGATCGGTTAC 822
Db 384 GlyLeuLysArgThrThrAspValMetPheGlyGlyLysValValValValCysGlyTyr 403
QY 823 GCGGATGTCGAGGAGGCTCGCTCAATCCCTCGTGGCCAGGCGCTCGGTTATCATC 882
Db 404 GlyGluValGlyLysGlyCysCysAlaAlaLeuLysAlaLeuValIleValIle 423
QY 883 ACAGAACTCGACCAATCTCGCTCTCCAGGCTGCCATGGAAGGCTACAGGTCGCGCGC 942
Db 424 ThrGluIleAspProIleCysAlaLeuGlnAlaCysMetAspGlyPheArgValValLys 443
QY 943 ATCCAGGAGTCTGCAAGGATGTCGATATCTTCTGTTACATGACACAGGAATTCGATATC 1002
Db ----- 1002
```



```
QY 1303 AAGAGTTTACACACTTCGAGAGCATCTCGATGAAGAGTGCCTCGCTCCACCTCGGA 1362
D 1400 LysLysValTyrValPheProLysHisLeuGluLysValAlaLeuHisLeuVal 419
QY 1363 TCTCTCGATGTCACCTCAAGAGCTTACAGAGAGGAGGCTGACTACATCACTGTC 1422
D 420 LysLeuGlyAlaArgLeuThrLysLeuSerArgCysThrLeuLeuCysThrAspAspPro 439
QY 1423 GTTAGGGT 1431
D 440 ValGluGly 442
RESULT 10
US-10-040-895-14
; Sequence 14, Application US/10040895
; Publication No. US20030162219A1
; GENERAL INFORMATION:
; APPLICANT: Sem, Daniel S.
; APPLICANT: Baker, Brian
; APPLICANT: Hansen, Mark R.
; TITLE OF INVENTION: Methods for Predicting Functional and
; TITLE OF INVENTION: Structural Properties of Polypeptides Using Sequence Models
; FILE REFERENCE: P-TB 5072
; CURRENT APPLICATION NUMBER: US/10/040,895
; CURRENT FILING DATE: 2002-04-09
; PRIOR APPLICATION NUMBER: US 09/753,020
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Rattus sp.
US-10-040-895-14
Alignment Scores:
Pred. No.: 2,46e-36 Length: 170
Score: 551.50 Matches: 112
Percent Similarity: 76.61% Conservative: 19
Best Local Similarity: 65.50% Mismatches: 39
Query Match: 20.15% Indels: 1
DB: 12 Gaps: 1
US-09-759-990A-1 (1-1461) x US-10-040-895-14 (1-170)
QY 721 AGTTTCATACATACGCTCGCGCCACTCCCTTATCGATGTATCAACGCTGCTCC 780
D 1 LysPheAspAsnLeuTyrGlyCysArgGluSerLeuLeuAspGlyIleLysArgAlaThr 20
QY 781 GATGTCATGATCGCGCGCAAGACAGCTCTCGTCATGGTTACGGCGATGTCGCAAGGC 840
D 21 AspValMetIleAlaGlyLysValAlaValAlaGlyTyrGlyAspValGlyLysGly 40
QY 841 TGGCTTCATCCCTCGCGCAAGCGCTCGCTTATCATCAAGACTCCACCAATC 900
D 41 CysAlaGlnAlaLeuArgGlyPheGlyAlaArgValIleIleThrGluLeuAspProIle 60
QY 901 TGGCTTCAGGCTCGCATCGAAGCTACAGCTCCGCGCATCGAGAACTCTCAAG 960
D 61 AsnAlaLeuGlnAlaAlaMetGluGlyTyrGluValThrThrMetAspGluAlaCysLys 80
QY 961 GATGTCATATCTCGTTATCATGCACAGGAACTCCGATATCATCTCTGTCATGATG 1020
D 81 GluGlyAsnIlePheValThrThrGlyCysValAspIleLeuLeuGlyArgHisPhe 100
QY 1021 GCCAGATCAAGATTAAGCTTATGTCGTATCATCGCCACTTCGATACCAAGATTCAT 1080
D 101 GluGlnMetLysAspAspAlaIleValCysAsnIleGlyHisPheAspValGluLeuAsp 120
QY 1081 ACAGATGGCTCATGAATACCCAGGCATCAGCACATCCCAATCAAGCCAGCAATACAC 1140
D 121 ValLysTrpLeuAsnGlu---AsnAlaValGluLysValAsnIleLysProGlnValAsp 139
```

```
QY 1141 ARGTGGAATTCACAGATGCCACCCCTATCTCTTCTTGTGAGGGCGGCTTCTTAAC 1200
D 140 ArgTyrLeuLeuLysAsnGlyHisArgIleLeuLeuAlaGlyArgLeuValAsn 159
QY 1201 CTTCGCTCGCTACAGGTCAACCATCTTTCGTT 1233
D 160 LeuGlyCysAlaMetGlyHisProSerPheVal 170
RESULT 11
US-09-738-626-4337
; Sequence 4337, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKKO
; APPLICANT: SENO, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4337
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4337
Alignment Scores:
Pred. No.: 2,02e-35 Length: 236
Score: 541.00 Matches: 130
Percent Similarity: 61.83% Conservative: 19
Best Local Similarity: 53.94% Mismatches: 84
Query Match: 20.66% Indels: 8
DB: 10 Gaps: 3
US-09-759-990A-1 (1-1461) x US-09-738-626-4337 (1-236)
QY 782 TCGAGACGAGCTGTGATACATCGATAGGAGTGGCGGACCGATGATTTATCGAAC 723
D 4 SerValAlaArgLeuMetProSerIleArgGluTrpArgValProTyrLeuLeuSerAsn 23
QY 722 TTGGACTTTGTACAGCGCTGATGATGCTGGGAGAGGAGTTTCCCTCTTC 663
D 24 LeuAspLeuValThrAlaSerLeuThrPheIleAlaGlyLysCysThrProSerSer 43
QY 662 TCGAGCTGGTAGAGCGGTGGACCTCTGTGTGTCTCTTCGGAACACCGCTTCATCGCG 603
D 44 AlaLysTrpTyrArgCysThrProValValIleSerSerValThrProLeuThrAla 63
QY 602 CGACCACTGTGTCAGCTGTTCTGTCTGTGGTGAAGACTGCTGTGATGTAGCAGA 543
D 64 SerAlaIleLeuProHisLeuProGlySerAlaAlaArgThrSerArgSerMetProLys 83
QY 542 ACGCAGCGGATTCGAGGTTGTGCTCTGCTGGTCTGGAAACAGCACCGCTGTTTCG 483
D 84 AsnAlaMetTyrSerSerGluSerLeuIleSerAlaGlyThrArgProAlaCysSer 103
QY 482 AATTCGAAGCCTTGGAGATGAGGAGTGATAGCATCACCACCATCATCGAACCTGCTGT 423
```

Db 104 TyrSerArgProArgIleThrAlaMetValAlaSerProProSerSerArgIleMetPhe 123  
QY 422 GGGCCCTTGACCATCTGGCCATGTGAGAGCGGTATGTGTCTCCAGTATCTGGAGT 363  
Db 124 GlySer--SerSerProGlnLeuLysIleThrPleuMetHisHisGlnTyrSerSer 142  
QY 362 GTTTCGCCCTTCAGGCGAGAGCTGGGATACCGGTGCTCTCTGTGTGCTGTGGG 303  
Db 143 AspSerProPheHisAlaAsnThrGlyThrProAlaGlySerSerThrValPro--Glu 161  
QY 302 CCACACACATAGCAGCAGCGGTGTATCTGTGTAGAGAAGATGTGACGAAGCCCAT 243  
Db 162 ProThrThrIleAlaAlaAlaSerSerTrpValGluLysMetLeuGlnGluAlaGln 181  
QY 242 CTGACATCAGCCACGAGCTGTGAGTCTCTCATGATGAGAGCGGTGTGCTGTGATG 183  
Db 182 ArgThrSerAlaProLysAlaValArgValSerIleSerThrAlaValTrpThrValMet 201  
QY 182 TGAGAGGAACACAGATCTGCACACCTTCATGCTTAGAAGCGGAATAACGCTCAGA 123  
Db 202 TrpIleGluProAlaIleArgAlaProPheLysGlyCysSerSerAlaAsnSerLeuArg 221  
QY 122 AGAACCATAGCTGGCATCTCTCTCAGCAAGGTAAGTCTTACGCGCGAACA 63  
Db 222 AsnCysMetArgProGlyIleSerTyrSerAlaArg-----ArgIle 235  
QY 62 TGG 60  
Db 236 Trp 236

## RESULT 12

US-09-738-626-4338  
; Sequence 4338, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIALI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738, 626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4338  
; LENGTH: 240  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4338

## Alignment Scores:

Pred. No.:	4,226-26	Length:	240
Score:	426.50	Matches:	98
Percent Similarity:	56.15%	Conservative:	39
Best local Similarity:	40.16%	Mismatches:	90
Query Match:	16.29%	Indels:	17
DB:	10	Gaps:	4

US-09-759-990A-1 (1-1461) x US-09-738-626-4338 (1-240)

QY 1080 ATCAATTCGTTATCGAAGTGGCCGATGTTACCGACAATAGCCCTTATCTTATCTGCGC 1021  
Db 3 IleAspLeuIleLysValThrAspValAlaGlnGlnSerValIleLeuHisLeuGlu 22  
QY 1020 CATCATGTCACACAGATATATCGACGTTTCCTGTGTCATGTAACGAGATATCGACATC 961  
Db 23 HisLeuLeuGluGlyLeuAspValLeuValAlaGlyArgGlyAspHisAspValGlyVal 42  
QY 960 CTTGACGACTTCTCCATGCGCGGACCTGTGTAGCTTCCATGCGAGCTCGAGAGCGCA 901  
Db 43 LeuAspGlyLeuIleAsnGlyAspHisArgValAlaIleHisGlnSerLeuLysSerVal 62  
QY 900 GATTGGTTCGAGTCTGTGATGATACCGGACGCCCTTGGCCACGGAGGATATGAGCGCA 841  
Db 63 AspTrpValSerPheGlyAspLeuAspAlaSerAlaLeuAlaValGlnSerLeuSerAla 82  
QY 840 GCCCTTGGCCGACATCGCGCTAAACCCATGACGAGCTGTCTGCGCGCGATCATGATC 781  
Db 83 AlaLeuAlaAspIleAlaValThrAlaAspLysHisValLeuAlaAlaHisGlnHisVal 102  
QY 780 GGAACACGGTTGATACCATGATTAAGGAGTGGCGGACCGGTAGATGTTATCGAACTT 721  
Db 103 SerGlyAlaValAspAlaValAspGlnGlyValAlaGlyAlaValLeuValIleLysLeu 122  
QY 720 GACCTTTCACACGCTGCTGACGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 661  
Db 123 GlyLeuGlyAspSerValValAspValHisArgTrpGlnArgGlnHisAlaPhePheSer 142  
QY 660 GAGCTGTGTAGAGCGGTGGACACCTGTGTGTCTTTCGAAACACCGCTTCATCGCGGC 601  
Db 143 GluValValGlnAlaValHisThrGlyGlyPheLeuGlyAspThrLeuAsnGlyLeu 162  
QY 600 ACCTACTGTGTGCGAGTGGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 541  
Db 163 SerAspLeuAlaProLeuAlaArgLeu-----CysSerLysAsn 175  
QY 540 GCA-----GCGGTATTTCGAGGTGTGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 505  
Db 176 LeuThrGlnHisAlaGlnGluCysAspValLeuLeuGlyIleValGlyLeuCysTrp--- 194  
QY 504 TCGAACACACCGCGCTGTTTCGAATTCGAAGCCCTTCGAGATGAGGAGTGTAGCATCACC 445  
Db 195 TrpAsnGlnThrSerLeuLeuValPheAlaThrAlaAsnAsnSerHisGlyGlyValThr 214  
QY 444 ACCTATCATCGACACACCTGCTGTGGCGCTTGACCATGTGGCCATGTGAGAGCGCGGTATGT 385  
Db 215 AlaValValGluAspHisValTrpGlnLeu---IleAlaProAlaGluAspLeuValAsp 233  
QY 384 GTTCTCCCGAGTA 373  
Db 234 AlaProProVal 237  
RESULT 13  
US-09-799-946-7  
; Sequence 7, Application US/09799946  
; Patent No. US2002009028A1  
; GENERAL INFORMATION:  
; APPLICANT: Gudkov, Andrei  
; Kazarov, Alexander  
; Mazo, Ilya  
; Robinson, Igor B  
; TITLE OF INVENTION: Methods for Identifying Genetic  
; Suppressor Elements and Genes Associated with Malignant  
; Growth in Cancer Cells  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Alligretti & Witcoff, Ltd.  
; STREET: 10 S. Wacker Drive, Suite 3000  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:

```

ZIP: 60606
COMPUTER READABLE FILE:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/799,946
FILING DATE: 06-Mar-2001
CLASSIFICATION: <Unknown>
PRIORITY INFORMATION:
APPLICATION DATA:
APPLICATION NUMBER: US/08/204,740
FILING DATE: 04-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: No. US20020099028AInan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 93,354-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
TELEX: 910-221-5317
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 9:

```

Alignment Scores:	
Pred. No.:	1.29e-13
Score:	271.00
Percent Similarity:	80.26%
Best Local Similarity:	69.74%
Query Match:	9.90%
DB:	9
Gaps:	0
Indels:	0
Mismatches:	15
Conservative:	8
Matches:	53
Length:	95

US-09-759-990A-1 (1-1461) x US-09-799-946-9 (1-95)

Qy	37	TACAGAAATGCCGACATCAACTCCATTCTCGGCCGTAAAGGAACTTACCCCTTCTCAG	96
		: : : : :       : : : : :       : : : : :	
Db	20	TyrLysValAlaAspIleGlyLeuAlaTrpGlyArgLysAlaLeuAspIleAlaGlu	39
		: : : : :       : : : : :       : : : : :	
Qy	97	AAGGAAATGCCAGGTCTTATGGTCTTCGTAGCGGTATTCGCTTCTTAAGCATTTGAAG	156
		: : : : :       : : : : :       : : : : :	
Db	40	AsnGluMetProGlyLeuMetArgMetArgGluMetTyrSerAlaSerLysProLeuLys	59
		: : : : :       : : : : :       : : : : :	
Qy	157	GGTGTCAGAATCTCGTGTCCCTCCACATCACAGTCCAGACAGCCGTCCTCATCCAGACA	216
		: : : : :       : : : : :       : : : : :	
Db	60	GlyAlaArgGlyAlaGlyCysLeuArgMetThrValGluThrAlaValLeuIleGluThr	79
		: : : : :       : : : : :       : : : : :	
Qy	217	CTCAGAGTCTTGTTGTGATGTACAGATGGGCTTCCTGCACATCTTC	264
		: : : : :       : : : : :       : : : : :	
Db	80	IvsValAlaLeuGlyValAlaGluAlaArgTrpSerCysAsnIlePhe	95
		: : : : :       : : : : :       : : : : :	

[illegible]

; ORGANISM: Staphylococcus epidermidis

US-10-172-502-4

Alignment Scores:

Pred. No.:	8.3e-09	Length:	2283
Score:	219.00	Matches:	111
Percent Similarity:	42.57%	Conservative:	81
Best Local Similarity:	24.61%	Mismatches:	167
Query Match:	8.00%	Indels:	92
DB:	12	Gaps:	13

US-09-759-990A-1 (1-1461) x US-10-172-502-4 (1-2283)

161	QY	TCGAANTCTCTGGTTCCCTCCAAATCAGACAGTCCAGACAGCCGCTCTCATCTCGACACATCA	220
162			
163			
164			
165			
166			
167			
168			
169			
170			
171			
172			
173			
174			
175			
176			
177			
178			
179			
180			
181			
182			
183			
184			
185			
186			
187			
188			
189			
190			
191			
192			
193			
194			
195			
196			
197			
198			
199			
200			
201			
202			
203			
204			
205			
206			
207			
208			
209			
210			
211			
212			
213			
214			
215			
216			
217			
218			
219			
220			
221	QY	CAGCTCTTGCTCATGCTCAGATGGGCTCTCTGCAACATCTTCTCTACACAAGATACAG	280
222			
223			
224			
225			
226			
227			
228			
229			
230			
231			
232			
233			
234			
235			
236			
237			
238			
239			
240			
241			
242			
243			
244			
245			
246			
247			
248			
249			
250			
251			
252			
253			
254			
255			
256			
257			
258			
259			
260			
261			
262			
263			
264			
265			
266			
267			
268			
269			
270			
271			
272			
273			
274			
275			
276			
277			
278			
279			
280			
281	QY	CCGCTCTCTCTATCGTTGTCGGCCCAA	307
282			
283			
284			
285			
286			
287			
288			
289			
290			
291			
292			
293			
294			
295			
296			
297			
298			



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: October 17, 2003, 20:07:15 ; Search time 5463 Seconds  
(without alignments)  
10940.685 Million cell updates/sec

Title: US-09-759-990A-1  
Perfect score: 1461  
Sequence: 1 atggctgcaaatcaactac.....ctggtcttaccgtttattaa 1461

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: en.ba.\*
- 16: en.fun.\*
- 17: en.hum.\*
- 18: en.in.\*
- 19: en.mu.\*
- 20: en.om.\*
- 21: en.or.\*
- 22: en.ov.\*
- 23: en.pat.\*
- 24: en.ph.\*
- 25: en.pl.\*
- 26: en.ro.\*
- 27: en.sts.\*
- 28: en.un.\*
- 29: en.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sv.\*
- 39: em.higo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1459.4	99.9	1461	6	AX195549	AX195549 Sequence
2	1440.2	98.6	1380	3	TVU40872	U40872 Trichomonas
3	655.2	44.8	313800	1	SC0939114	AL939114 Streptomy
4	642.4	44.0	30425	1	AP005041	AP005041 Streptomy
5	615.2	42.1	3800	1	SAR416377	AJ416377 Streptomy
6	596.4	40.8	1557	6	AX063937	AX063937 Sequence
7	596.4	40.8	1557	6	AX244105	AX244105 Sequence
8	596.4	40.8	1939	6	AX397890	AX397890 Sequence
9	596.4	40.8	332050	1	AP005276	AP005276 Corynebac
10	596.4	40.8	349980	6	AX127145	AX127145 Sequence
11	594	40.7	1422	6	AX120920	AX120920 Sequence
12	594	40.7	1422	6	BD163037	BD163037 Novel pol
13	567.2	38.8	308750	1	AP005216	AP005216 Corynebac
14	563.8	38.6	210614	1	AB088224	AB088224 Streptomy
15	560	38.3	12968	1	AE012843	AE012843 Chlorobiu
16	558.4	38.2	135638	1	AF484556	AF484556 Streptomy
17	552.6	37.8	9427	1	AE011711	AE011711 Xanthomon
18	541.2	37.0	9054	1	AE012174	AE012174 Xanthomon
19	536	36.7	1458	8	AY224544	AY224544 Oryza sat
20	535.4	36.6	1488	1	AF262755	AF262755 Mycobacte
21	535.4	36.6	1488	6	AX023852	AX023852 Sequence
22	535.4	36.6	18249	1	AE007145	AE007145 Mycobacte
23	535.4	36.6	36330	1	MTX20811	Z95121 Mycobacteri
24	535.4	36.6	308050	1	EX248345	EX248345 Mycobacte
25	529.6	36.2	1396	6	AX063941	AX063941 Sequence
26	529.6	36.2	1396	6	AX244109	AX244109 Sequence
27	529.6	36.2	1708	8	WHFSSH	L11872 Triticum ae
28	512.6	35.1	1837	8	MCU79766	U79766 Mesembryant
29	511.4	35.0	312050	1	MLSPRTN3	AL583919 Mycobacte
30	508.2	34.8	1727	8	AY085669	AY085669 Arabidops
31	506.6	34.7	1458	8	AY090284	AY090284 Arabidops
32	506.6	34.7	1626	8	AY081468	AY081468 Arabidops
33	506.6	34.7	1729	8	AY042866	AY042866 Arabidops
34	506.6	34.7	1802	8	AF059581	AF059581 Arabidops
35	506.6	34.7	1821	8	AF325037	AF325037 Arabidops
36	506.6	34.7	1821	8	AY049279	AY049279 Arabidops
37	506.6	34.7	1823	8	BT002404	BT002404 Arabidops
38	506	34.6	13407	1	AE008977	AE008977 Agrobacte
39	506	34.6	13452	1	AE007946	AE007946 Agrobacte
40	501.8	34.3	1798	8	PSGADHY	X79905 Phalaenopsi
41	500.6	34.3	305961	1	AE016937	AE016937 Bacteroid
42	499.6	34.2	260050	1	SME591782	AL591782 Sinorhizo
43	497	34.0	1790	8	ATSADLHH	Z97059 Arabidopsis
44	490	33.5	1812	6	AR071283	AR071283 Sequence
45	490	33.5	1812	6	E03902	E03902 A gene invo

ALIGNMENTS

RESULT 1	AX195549	AX195549	1461 bp	DNA	linear	PAT 28-AUG-2001
LOCUS	Sequence 1 from Patent WO0151651.					
DEFINITION	AX195549					
ACCESSION	AX195549.1	GI:15386074				
VERSION	AX195549.1					
KEYWORDS	Trichomonas vaginalis					
SOURCE	Trichomonas vaginalis					
ORGANISM	Eukaryota; Parabasalida; Trichomonadida; Trichomonadidae; Trichomonas;					
REFERENCE	1					
AUTHORS	Hoffman, R.M., Xu, M. and Han, Q.					
TITLE	High expression and production of high-specific activity recombinant s-adenosyl homocysteinease (sahh) and improved assays					

for s-adenosylmethionine (sam)  
Patent: WO 0151651-A 1 19-JUL-2001;  
Anticancer, Inc. (US)

JOURNAL  
FEATURES  
source  
Location/Qualifiers

1..1461  
/organism="Trichomonas vaginalis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:5722"

BASE COUNT 372 a 427 c 335 g 327 t

ORIGIN

Query Match 99.9%; Score 1459.4; DB 6; Length 1461;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1460; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGCTTGCAATCACCTACTGTGCTCCATTGCGAGTACAGAAATGCGACATCAACCTC 60  
DB 1 ATGGCTTGCAATCACCTACTGTGCTCCATTGCGAGTACAGAAATGCGACATCAACCTC 60  
QY 61 CATGTTCTCGCGCTTAAGGAACCTTACCTTGCTGAGAGGAATGCGAGGCTTTATGTT 120  
DB 61 CATGTTCTCGCGCTTAAGGAACCTTACCTTGCTGAGAGGAATGCGAGGCTTTATGTT 120  
QY 121 CTTGCTGAGCGTATTCCGCTTCTTAAGCCATTGAGGGTGTGCAATCTCTGTTCCCTC 180  
DB 121 CTTGCTGAGCGTATTCCGCTTCTTAAGCCATTGAGGGTGTGCAATCTCTGTTCCCTC 180  
QY 181 CACATGACAGTCCAGACAGCGCTCTCATCGAGACACTCACAGCTTTGCTGCTGATGTC 240  
DB 181 CACATGACAGTCCAGACAGCGCTCTCATCGAGACACTCACAGCTTTGCTGCTGATGTC 240  
QY 241 AGATGGGCTTCTGCAACATCTTCTTCAACAAGATACAGCGCTGCTGCTATCTGTTGTC 300  
DB 241 AGATGGGCTTCTGCAACATCTTCTTCAACAAGATACAGCGCTGCTGCTATCTGTTGTC 300  
QY 301 GGGCCAAAGGACACAGAGAGAGCGGCTATCCAGGCTTCTGCTGGAAGGGCGAA 360  
DB 301 GGGCCAAAGGACACAGAGAGAGCGGCTATCCAGGCTTCTGCTGGAAGGGCGAA 360  
QY 361 ACATCCCAAGATATCGGAGAACACATACCGCGCTCTCATGATGGCCAGATGTCAGGC 420  
DB 361 ACATCCCAAGATATCGGAGAACACATACCGCGCTCTCATGATGGCCAGATGTCAGGC 420  
QY 421 CCAGCAGGCTTGCATGATGTTGTTGATGCTCACTCTCATCTCCAGGCTTCGAA 480  
DB 421 CCAGCAGGCTTGCATGATGTTGTTGATGCTCACTCTCATCTCCAGGCTTCGAA 480  
QY 481 TTGAAACAGCGGCTGCTTCCAGAGCCAAACAGAGCTGACAACTCGAATACCGCTGC 540  
DB 481 TTGAAACAGCGGCTGCTTCCAGAGCCAAACAGAGCTGACAACTCGAATACCGCTGC 540  
QY 541 GTTCTGCTACATCAGCAGGCTTCCAAACAGAGAACCACTGGCAGCAGTTGCT 600  
DB 541 GTTCTGCTACATCAGCAGGCTTCCAAACAGAGAACCACTGGCAGCAGTTGCT 600  
QY 601 GCGGCGATGAAGGTTTTCGAGAGACAAACAGAGTGTCCACGCTCTTACAGCTC 660  
DB 601 GCGGCGATGAAGGTTTTCGAGAGACAAACAGAGTGTCCACGCTCTTACAGCTC 660  
QY 661 GAGAAGGAGGCAAACTCTTCCAGGCTATCAACGCTGTCAGAGCTGTTCAGAGTCC 720  
DB 661 GAGAAGGAGGCAAACTCTTCCAGGCTATCAACGCTGTCAGAGCTGTTCAGAGTCC 720  
QY 721 AAGTTGCATTAATCAGGCTGCGGCTACTCTTATCGATGATGATCAACGCTGCTCC 780  
DB 721 AAGTTGCATTAATCAGGCTGCGGCTACTCTTATCGATGATGATCAACGCTGCTCC 780  
QY 781 GATGTCATCATCGGCGGCAAGACAGCTCTGCTCATGGGTTACGGGATGTCGCAAGGC 840  
DB 781 GATGTCATCATCGGCGGCAAGACAGCTCTGCTCATGGGTTACGGGATGTCGCAAGGC 840  
QY 841 TGGCGCTCAATCCCTCGGTGGCCAGGCGCTCGCGTTATCATCAGAACTCGACCCAAATC 900

DB 841 TGGCGCTCAATCCCTCGGTGGCCAGGCGCTCGCGTTATCATCAGAAAGTCCAGCCCAATC 900  
QY 901 TGGCGCTTCCAGGCTGCCATGGAAGGCTTACAGGTTCGGCGCATCGAGGAAGTGGTCAAG 960  
DB 901 TGGCGCTTCCAGGCTGCCATGGAAGGCTTACAGGTTCGGCGCATCGAGGAAGTGGTCAAG 960  
QY 961 GATGTCGATATCTTCTTACATGACAGGAACCTGCGATATCATCTCTGTTGACATGATG 1020  
DB 961 GATGTCGATATCTTCTTACATGACAGGAACCTGCGATATCATCTCTGTTGACATGATG 1020  
QY 1021 CCCAGATGAGGATAAGGCTATTGTCGTAACATCGGCACTTCGATAACGAAATGAT 1080  
DB 1021 CCCAGATGAGGATAAGGCTATTGTCGTAACATCGGCACTTCGATAACGAAATGAT 1080  
QY 1081 ACAGATGGCTCTCATGAAATACCCAGGCATCAAGCACATCCCAATCAAGCCAGAAATGAC 1140  
DB 1081 ACAGATGGCTCTCATGAAATACCCAGGCATCAAGCACATCCCAATCAAGCCAGAAATGAC 1140  
QY 1141 ATGTGGGAATTCAGATGGCCAGCTATCTCTGTTGCTGAGGCGGCTTCTTAAAC 1200  
DB 1141 ATGTGGGAATTCAGATGGCCAGCTATCTCTGTTGCTGAGGCGGCTTCTTAAAC 1200  
QY 1201 CTTGCTGCGCTACAGTCAACCATCTTCTGTTATGTCATTCATTCACAAACGACACA 1260  
DB 1201 CTTGCTGCGCTACAGTCAACCATCTTCTGTTATGTCATTCATTCACAAACGACACA 1260  
QY 1261 CTGCTCAGCTCGACCTCTTACGAAAGAGAGAAATCTCGAAGAGAGGTTTACACACTT 1320  
DB 1261 CTGCTCAGCTCGACCTCTTACGAAAGAGAGAAATCTCGAAGAGAGGTTTACACACTT 1320  
QY 1321 CCGAAGCATCTCGATGAGAAAGTCTGCTGCTCCACCTCGGATCTCTCGATGTCACCTT 1380  
DB 1321 CCGAAGCATCTCGATGAGAAAGTCTGCTGCTCCACCTCGGATCTCTCGATGTCACCTT 1380  
QY 1381 ACAAGCTTACAGAGAGGAGGCTGACTACATCAAGTTCCAGTTGAGGGTCTTACAAG 1440  
DB 1381 ACAAGCTTACAGAGAGGAGGCTGACTACATCAAGTTCCAGTTGAGGGTCTTACAAG 1440  
QY 1441 TCTGATGCTTACCGTTATTAA 1461  
DB 1441 TCTGATGCTTACCGTTATTAA 1461

RESULT 2

TVU40872 1882 bp DNA linear INV 31-OCT-1996  
LOCUS Trichomonas vaginalis S-adenosyl-L-homocysteine hydrolase gene,  
complete cds.  
DEFINITION U40872 GI:1118008  
ACCESSION U40872  
VERSION U40872.1  
KEYWORDS Trichomonas vaginalis  
SOURCE Trichomonas vaginalis  
ORGANISM Trichomonas vaginalis  
REFERENCE 1 (bases 1 to 1882)  
AUTHORS Bagnara, A.S., Tucker, V.E., Minotto, L., Howes, E.R., Ko, G.A.,  
Edwards, M.R., and Dawes, I.W.  
TITLE Molecular characterization of adenosylhomocysteine from  
Trichomonas vaginalis  
Mol. Biochem. Parasitol. 81 (1), 1-11 (1996)  
JOURNAL 97047381  
MEDLINE 8892301  
PUBMED 8892301  
REFERENCE 2 (bases 1 to 1882)  
AUTHORS Bagnara, A.S.  
TITLE Direct Submission  
JOURNAL Submitted (19-NOV-1995) Aldo S. Bagnara, Biochemistry and Molecular  
Genetics, The University of New South Wales, Sydney, New South  
Wales, 2052, Australia  
FEATURES Location/Qualifiers  
1..1882  
/organism="Trichomonas vaginalis"  
/mol\_type="genomic DNA"

```
/strain="WAA38"
/specific host="Homo sapiens"
/db xref="taxon:5722"
1. 290
291..1751
/EC_number="3.3.1.1"
/codon_start=1
/product="S-adenosyl-L-homocysteine hydrolase"
/protein_id="AAC47319.1"
/db xref="GI:1118009"
/translation="VACKSPAGAPFEVRIADINLHVLGRKELTLAEKEMPLMVLIER
YSASKPKGVRLSGSLEMTVQAVLIETLALGADVWASCNIPSTODTAAAIIVGP
TETEPAGPEFPANKEITLPEWPDGQPPQVDDGGDALLLSKGFE
FGTEKADLELXRCVIALTKQFNQDKHMHVTAAGMNGVSEETTTGVHRLY
QLEKGGKLLFPAINVNDVATKSFEDNIYGRHSLIDGINRASDVIMGKFTALVWGVD
VKGKGAQSLRGQGVRIITEVDPICALQAWMEGYQVRIIEEVKVDVIFVTCNGNDI
ISVDMQMKDRAIVGNIGHPDNEIDTGLMKVPGIKHPIKPEYDMWEPDPGHALL
LAERGLNLGCAIHTGHPFSVMSFTNOTLAOLDLVEKGNLEMKVYIPLKHLDEEVR
LHGLSLDVHLTKLTQKQADYINVPVEGPKSDAIR"
1752..1882
3'UTR
polyA_signal
BASE COUNT 506 a 499 c 395 g 482 t
ORIGIN
Query Match 98.6%; Score 1440.2; DB 3; Length 1882;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1448; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
Qy 1 ATGGCTTCCAATCACCTACTGTGCTCATTCGAGTACAGAAATTCGCGACATCAACCTC 60
Db 291 ATGGCTTCCAATCACCTGCTGTGCTCCATTCGAGTACAGAAATTCGCGACATCAACCTC 350
Qy 61 CATGTTTCGGCGCTAGGAACTTACCTTCGCTGAGAGGAATGCGAGGTCTTATGGTT 120
Db 351 CATGTTTCGGCGCTAGGAACTTACCTTCGCTGAGAGGAATGCGAGGTCTTATGGTT 410
Qy 121 CTTGCTGAGCGTTTATCCGCTCTAAGCCATTGAAGGTGTCAGAACTCTGCTTCCTC 180
Db 411 CTTGCTGAGCGTTTATCCGCTCTAAGCCATTGAAGGTGTCAGAACTCTGCTTCCTC 470
Qy 181 CACATGACAGTCCAGACAGCGGTCCTCATGAGACACTCAACGCTTTGCTGCTGATGTC 240
Db 471 CACATGACAGTCCAGACAGCGGTCCTCATGAGACACTCAACGCTTTGCTGCTGATGTC 530
Qy 241 AGATGGGTTCTCTCAACATCTTCTACACAGATACAGCGCTGCTGCTATGTTGTC 300
Db 531 AGATGGGTTCTCTCAACATCTTCTACACAGATACAGCGCTGCTGCTATGTTGTC 590
Qy 301 GGCCCAACAGGCACACAGAGAGCCAGCCGCTATCCAGTCTTCGCTTGGAAAGGGGAA 360
Db 591 GGCCCAACAGGCACACAGAGAGCCAGCCGCTATCCAGTCTTCGCTTGGAAAGGGGAA 650
Qy 361 ACATCCGAGAACTGGGAGAACATACCGCGCTTCAATGCGCCAGATGGTCAAGGC 420
Db 651 ACATCCGAGAACTGGGAGAACATACCGCGCTTCAATGCGCCAGATGGTCAAGGC 710
Qy 421 CCACAGAGGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
Db 711 CCACAGAGGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 770
Qy 481 TTCCGAAACAGCGCGTGTGTTCCAGAGCCACACAGAGGCTGACAACTCCGAATACCGTGC 540
Db 771 TTCCGAAACAGCGCGTGTGTTCCAGAGCCACACAGAGGCTGACAACTCCGAATACCGTGC 830
Qy 541 GTTCTTGCTACACTCAAGCAGGTTCCTCAACCAAGCAAGAACCACTGGGCAACAGTTGCT 600
Db 831 GTTCTTGCTACACTCAAGCAGGTTCCTCAACCAAGCAAGAACCACTGGGCAACAGTTGCT 890
Qy 601 GCGGCGATGAACGGTGTTCGAGAGAGACACACAGGTGTCACCGCTCTTACCAAGTTC 660
Db 891 GCGGCGATGAACGGTGTTCGAGAGAGACACACAGGTGTCACCGCTCTTACCAAGTTC 950
Qy 661 GAGAAGGAGGGGAAACTCTCTTCCAGCCATCAACGCTCAACGACGCTGTTACAAAGTTC 720
```

```
Db 951 GAGAAGGAGGGGAAACTCTCTTCCAGCCATCAACGCTCAACGAGCTGTGTACAAAGTCC 1010
Qy 721 AAGTTTGGATAAATCTACGGCTGCGGCACCTCCCTTATCGATGGTATCAACCGTGTCC 780
Db 1011 AAGTTTGGATAAATCTACGGCTGCGGCACCTCCCTTATCGATGGTATCAACCGTGTCC 1070
Qy 781 GATCTCATGATGGCGGCAAGACAGCTCTCTGTCATGGTTACGGCGATGTGCGCAAGGCG 840
Db 1071 GATCTCATGATGGCGGCAAGACAGCTCTCTGTCATGGTTACGGCGATGTGCGCAAGGCG 1130
Qy 841 TCGCCTCAATCCCTCCGTCGCAAGGCGCTCGGTTATCATCAGAACTCGACCAATC 900
Db 1131 TCGCCTCAATCCCTCCGTCGCAAGGCGCTCGGTTATCATCAGAACTCGACCAATC 1190
Qy 901 TCGCCTCTCCAGCTGCCATGGAAGGCTACAGGTCGCGCGCATCGAGGAAGTCTGCAAG 960
Db 1191 TCGCCTCTCCAGCTGCCATGGAAGGCTACAGGTCGCGCGCATCGAGGAAGTCTGCAAG 1250
Qy 961 GATCTCATGATCTGTTTACATGCAAGGAACTCGGATATCATCTCTGTTGATGATG 1020
Db 1251 GATCTCATGATCTGTTTACATGCAAGGAACTCGGATATCATCTCTGTTGATGATG 1310
Qy 1021 GCCCAGATGAAGGATAAGGCTATTGTCGTAACATCGGCACTTCGATTAACGAAATGAT 1080
Db 1311 GCCCAGATGAAGGATAAGGCTATTGTCGTAACATCGGCACTTCGATTAACGAAATGAT 1370
Qy 1081 ACAGATGGCTCATGAAATACCCAGGATCAAGCATCCCAATCAAGCCAGAAATACGAC 1140
Db 1371 ACAGATGGCTCATGAAATACCCAGGATCAAGCATCCCAATCAAGCCAGAAATACGAC 1430
Qy 1141 ATGTGGGAATTCAGATGGGCAAGCTATCTCTCTTCCTGCTGAGGGCGGCTTCTTAAC 1200
Db 1431 ATGTGGGAATTCAGATGGGCAAGCTATCTCTCTTCCTGCTGAGGGCGGCTTCTTAAC 1490
Qy 1201 CTTGGCTGCGCTACAGGTCACCCATCTTTCGTTATGTCATGTCATTCACAAACCAAGACA 1260
Db 1491 CTTGGTGGCTACAGGTCACCCATCTTTCGTTATGTCATGTCATTCACAAACCAAGACA 1550
Qy 1261 CTCGCTCAGCTCAGCTCTPAGAAAAGAGAGGAAATCTCGAGAGAGAGGTTTACACAT 1320
Db 1551 CTCGCTCAGCTCAGCTCTPAGAAAAGAGAGGAAATCTCGAGATGAAGGTTTACACAT 1610
Qy 1321 CCGAAGCATCTCGATGAGAGAGGCTCGCTCGCTCCACTCGGATCTCTCGATGTCACCT 1380
Db 1611 CCGAAGCATCTCGATGAGAGAGGCTCGCTCGCTCCACTCGGATCTCTCGATGTCACCT 1670
Qy 1381 ACAAGCTTACACAGAGCGGCTGACTACATCAAGTTTCCAGTTGAGGGTCTCTTCAAG 1440
Db 1671 ACAAGCTTACACAGAGCGGCTGACTACATCAAGTTTCCAGTTGAGGGTCTCTTCAAG 1730
Qy 1441 TCTGATCTTACCGTTATTAA 1461
Db 1731 TCTGATCTTACCGTTATTAA 1751
RESULT 3
SC0939114/c 313800 bp DNA linear BCT 11-FEB-2003
LOCUS
DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 11/29.
ACCESSION AL391114 AL094852 AL136058 AL138851 AL138852 AL13641 AL163672
AL353832 AL353862 AL355774 AL42165 AL499570 AL596030 AL596138
AL645882
VERSION AL391114.1 GI:24413773
KEYWORDS
SOURCE Streptomyces coelicolor A3(2)
ORGANISM Streptomyces coelicolor A3(2)
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
1
AUTHORS Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L.,
Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H.,
Harper, D., Bateson, A., Brown, S., Chandra, G., Chen, C.W., Collins, M.,
```

Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S., Huang,C.H., Klesner,T., Larke,L., Murphy,L., Oliver,K., O'Neill,S., Rabinowitsch,B., Rajandream,M.A., Rutherford,K., Rutter,S., Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S., Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrelli,B.G., Parkhill,J. and Hopwood,D.A.

Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)

Nature 417 (6885), 141-147 (2002)

21996410

PUBMED 12000953

2 (bases 1 to 313800)

Bentley,S.D.

Direct Submission

Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sdbesanger.ac.uk

On or before Oct 26, 2002 this sequence version replaced

gi:5531349, gi:20520908, gi:6911971, gi:6912004, gi:20520853, gi:12224900, gi:20520885, gi:7649483, gi:7799193, gi:20520703, gi:14018350, gi:20520873, gi:14970931.

Location/Qualifiers

1..313800

/organism="Streptomyces coelicolor A3(2)"

/mol\_type="genomic DNA"

/strain="A3(2)"

/db\_xref="taxon:100226"

182..1027

/gene="SCO2781"

/note="synonym: SCC105.12"

182..1027

/gene="SCO2781"

/note="SCC105.12, hypothetical protein, len: 281 aa; similar to TR:CA85689 (EMBL:AL121596) Streptomyces coelicolor hypothetical 28.7 kD protein, 259 aa; fasta scores: opt: 261 z-score: 302.6 E(): 2e-09; 34.9% identity in 255 aa overlap"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="CA887218.1"

/db\_xref="GI:7544045"

/translation="MRVPDPAPPAGSKRHPSPALARVIFAGPDLRAFRSGLDQSLSLFLPHGQAEPAVPVPELGEKQWGRLEPDEVRAVMSVTLRSLRDYDGTATIDVDVFLHLEPDSGLOGAARAKADAGDGVLLGPVADNRAIRRPEDTDLYVINGDETAPACAIVELPAGTRARWLQVPHADVDQDLTADAEIWLVDGADGFATLATLRADLPAPHEPIYVIAGSGCVKSLRHRHFGVGRVDVRRRVTFVGYRRRGUTERQLEEQ"

1178..1183

1191..2633

/gene="SCO2782"

/note="synonym: SCC105.13"

1191..2633

/gene="SCO2782"

/note="SCC105.13, probable pyridoxal-dependent decarboxylase, len: 480 aa; similar to SW:RHSE RHIME (EMBL:AF110737) Rhizobium meliloti L-2,4-diaminobutyrate decarboxylase (EC 4.1.1.-) RhSB, 495 aa; fasta scores: opt: 1111 z-score: 1326.2 E(): 0; 41.9% identity in 458 aa overlap. Contains Pfam match to entry PF00282 pyridoxal\_dec, Pyridoxal-dependent decarboxylase conserved domain"

/codon\_start=1

/transl\_table=11

/product="putative pyridoxal-dependent decarboxylase"

/protein\_id="CA887219.1"

/db\_xref="GI:7544046"

/translation="MRSKHLNDTTAEQYRRSVTEGVERVAAKLAITTFPTGVTVDALSPRIADLDPLADYAAVLDEVDVLRDAPVPHPRYLAHNCFWVWPALGCAVL SAVNSLTDWQSGAGGTJLERLIDWTCARIGLGAADGVFTSGGTQSNQALLLAEREAKADFDADLEIFASASHSFKAKLGLGPDPAVVISIPVDRKMTVALARELERCARDGLVPMAVATGTTDFGSDLPETAGLCGQYGVGWHVDAAVGGCLASLYKRDRTGLERADSVTVNDYHKSFQPVSSSAVLVDAATLRHATYHAELNPRRWQERIPN

## misc\_feature

QVKSQTTTRRDALKLMTURVMGADGICVLDFEVCDLAAGWKLLAADPREDVVVVVPS-STLVFPHIPADVTDAEDIRANLYARKALFASGDVAVGTACVGRHYLKTLLNPETTPADIAVLDLIAGHAEQYLGIDSLDRAS"

1605..2399

/gene="SCO2782"

/note="Pfam match to entry PF00282 pyridoxal dec, Pyridoxal-dependent decarboxylase conserved domain, score 157.60, E-value 4.6e-45"

2656..3897

/gene="SCO2783"

/note="synonym: SCC105.14"

2656..3897

/gene="SCO2783"

/note="SCC105.14, probable monooxygenase, len: 413 aa; similar to SW:RHSE RHIME (EMBL:AF110737) Rhizobium meliloti rhizobactin siderophore biosynthesis protein Rhse (EC 1.-.-.-), 454 aa; fasta scores: opt: 616 z-score: 727.2 E(): 0; 42.3% identity in 430 aa overlap and to SW:LUOD\_ECOLI (EMBL:M18368) Escherichia coli L-lysine 6-monooxygenase (EC 1.13.12.10) LucD, 425 aa; fasta scores: opt: 802 z-score: 946.5 E(): 0; 31.6% identity in 414 aa overlap"

/codon\_start=1

/transl\_table=11

/product="putative monooxygenase"

/protein\_id="CA887220.1"

/db\_xref="GI:7544047"

/translation="NGIGLGFNGLACLTPEVALNGVFLFKSPDEFHAGMFLDGAHLQTPFMSDLVTIADTSPSYFNLYLKEQRLYSFVIRENFPYLRVYDDYCSWARKLSVRFSTVTVEYTDREELVAVATSGTFRARRLVLTGTPHPIDPACRAGLADFLHNSRVHRALSKVKKYSITLVGSGQSAEIVQDLISLIDVHGVLGNWVRSRPRFPLVYKTLNLETSPEYVYVHALPEYRVLTAQKGLGKLGIDGLINEIFLLYQKELGCPVPTLLNSALTSARYADSTYTLGRFQEQGTFEIETGLVLAIGYRYTEFLKPVDRDLVTRSGNFIDIGIRVAVVGTGGVFLQNGVHAHSVTSPLQMGAYRNSCLVRELGREYYPVEQSINAFQEFAY"

3881..3885

/gene="SCO2783"

3894..4448

/gene="SCO2784"

/note="synonym: SCC105.15"

3894..4448

/gene="SCO2784"

/note="SCC105.15, probable acetyltransferase, len: 184 aa; similar to SW:RHSD RHIME (EMBL:AF110737) Rhizobium meliloti rhizobactin siderophore biosynthesis protein Rhsd, 196 aa; fasta scores: opt: 257 z-score: 326.7 E(): 9.2e-11; 33.7% identity in 172 aa overlap and to TR:AA46626 (EMBL:AF162771) Pseudomonas aeruginosa aminoglycoside 6'-N-acetyltransferase Aac, 193 aa; fasta scores: opt: 175 z-score: 226.3 E(): 3.6e-05; 28.2% identity in 174 aa overlap"

/codon\_start=1

/transl\_table=11

/product="putative acetyltransferase"

/protein\_id="CA887221.1"

/db\_xref="GI:7544048"

/translation="NSRLSTTTPVGALTLPVDELTDVILLHGLMTHPKSAFWMMODAFLVDVFRAYMELAADSHQAHGLHGDGPATFETVDPARELVLGVEPFGVGNHFLVATDRPHVGTFRVAVITVMTLFDATRRVVVEFDVINTAVHALNAAGVFPEREIQKPERKALLSFCTREQFAKAVSA"

4445..6232

/gene="SCO2785"

/note="synonym: SCC105.16"

4445..6232

/gene="SCO2785"

/note="SCC105.16, hypothetical protein, len: 595 aa; similar to SW:RHSE RHIME (EMBL:AF110737) Rhizobium meliloti rhizobactin siderophore biosynthesis protein Rhse, 601 aa; fasta scores: opt: 1094 z-score: 1325.6 E(): 0; 32.4% identity in 592 aa overlap and to TR:P94255 (EMBL:U61153) Bordetella bronchiseptica Alcc, 618 aa; fasta scores: opt: 2037 z-score: 2470.2 E(): 0; 51.5% identity in 592 aa overlap"

## TITLE

JOURNAL  
MEDLINE  
PUBMED  
AUTHORS  
REFERENCES

## TITLE

JOURNAL

## COMMENT

## FEATURES

source

## gene

## CDS

## RBS

## gene

## CDS

## gene

## CDS

	codon_start=1	transl_table=11	product="conserved hypothetical protein"	protein_id="CAB87222.1"	db_xref="GI:7544049"	translation="MSIADAAVAHITPERWEANRLLVKALABFTHERLLITPERSPDD GGGQTVVSDSDGQAVRPFATVLRDQVDAASVTRHGDGAELP.AALDFEILKQ TLGSEIILPVILEISLSTGCTYKLTQPLSSAELARSGDFORVETGNTGHPGCFV ANNRGFGHIGRLGDPADYLLIPVHPQWNKLTVTFAAEAVAGHLVCLGEGDEGLAQ VDFHVLGRLGDPADYLLIPVHPQWNKLTVTFAAEAVAGHLVCLGEGDEGLAQ QSTRTFNAGPKGVKATLSVLANMGFRLGSAAYMEATPAINDMLARLIGDPLVK ETGLSITIRRAAGVGHLEQYQADSYSPKMLAALWRESPVPSIREGETLATNASL VHODGSEAFACALERSGLTPEWLRHYLRAVYVPLHSPYAVLIVMHPGENVILV LADGVVRVAYQDAEIRIATMDPAVLPEVSIADVDPDDKLLISFTDVPDCFRP LAANLABEIGVIEDPAWRTVAETREYQSEVPELADKFERIDMFAPEFALSCLNRQLQ RDNKQWIDADSPGALQLVGLTKPLAGR"	gene	complement(6302..7909)	names="SCO2786"	notes="SC0105.17c, hexA, beta-N-acetylhexosaminidase precursor, len: 535 aa; highly similar to TR:085361 (EMBL:AF063001) Streptomyces plicatus B-N-acetylhexosaminidase, Hex, 561 aa; fasta scores: opt: 2991 z-score: 3318.1 E(): 0; 93.2% identity in 469 aa overlap and to TR:BAAB8876 (EMBL:AB015350) Streptomyces thermoviolaceus beta-N-acetylglucosaminidase precursor NagB, 541 aa; fasta scores: opt: 2856 z-score: 3168.6 E(): 0; 76.7% identity in 537 aa overlap. Contains Pfam match to entry PF00728 Glyco_hydro_20, Glycosyl hydrolase family 20 and possible N-terminal region signal peptide sequence"	codon_start=1	transl_table=11	product="beta-N-acetylhexosaminidase"	protein_id="CAB87223.1"	db_xref="GI:7544050"	translation="MRPHRRHRTTPRIITELIGSLLLVAAVGMTTGAAPVRKAAAE TPIDRVIPASVEGGAPVRIITGRTHIRVDDSRBARRVGYLDLRPATGYRLFPV	Query Match	Best Local Similarity	Matches	Conservative	0;	Mismatches	473;	Indels	6;	Gaps	1;																									
QY	34	GAGTACAGAAATTCGACATCAACCTCCAGTTCCTGGCGGTATTCGCGCTTCTAAGCCATTG	153	272138	GACITCAAGATTCGCGACCTCTCCCTGGCGCGTTCGCGCAGAGATCACCTCGCC	272079	QY	94	GAGAAGAAATGCGAGTCTATGTCTTCGTGAGCGTTATTCGCTTCTAAGCCATTG	153	272078	GAGCAGAGATGCGCGCTCATGGCGATCCGCMAGGATGCGCGAGGCCACCGCCTC	272019	QY	154	AAGGTGTCAGAAATCTCTGGTTCCTCCACATGACAGTCCAGACGCGCTCTCATCGAG	213	272018	GCGGCGCGCGCTCACCGGCTCCTGTCACATGACGTGCAGACGCGCTCATCGAG	271959	QY	214	ACATCAGACTTTGGTCTGATGTCAGATGGCTTCTTGCAATCTTCTTACACAA	273	271958	ACCTGTGCGCTTCGGCGCGAGTTCGCTGGGCTCTGTGAACATTTCTCACCCAG	271899	QY	274	GATACGCGCTGCTGATTCGTTTCGGGCCAACAGGCACACAGAACCCAGCGGT	333	271898	GACCACGGCGCGCCCATCGCTGCGCCCGAAGCGCACGCCGACACCCGAGGC	271839	QY	334	ATCCAGTCTTTCGGCTGGAAAGGCGGAACATCTCCAGATATCTGGGAGAACACATACCGC	393	271838	GTCCCGCTTTCGCTTGGAAAGGCGGAGCCCTCGAGAGTACTGTGTGTGTCACCGAGCAG	271779	QY	394	GCTCTCAATGCGCCAGATGTCACAA-----GGCCCCACAGCAGTTCGTGATGATGGTGGT	447	271778	GCGCTGACCTGGCGAGACACCCCGCACCGGCGCCCGAACATGATCTCTGGACGACGCGGT	271719	QY	448	GATGCTACCTCTCATCTCCAGGGCTTCGAATTCGAACACGCGGCTGCTGTTCACAGAG	507

Db	271718	GA	CGCCACCTCTCTGTCACAAAGGGGTGGAGTACGAAAGACGGCAAGGTCCCTCG	271659
Qy	508	CCAACAGAACTCAACAACTCGAATACCGGTGGTCTTGTCTACACTCAAGCAGGTCTTC	567	
Db	271658	GTCCGACACCGCCGAGTCCGACGACGCGGTCACTCTCGAACTCTCACCGCACGGTC	271599	
Qy	568	AACCAAGACAAAGAACCACTGGGCACACAGTTGTCTGCGCGCATGAACGGTGTTCGAAAGAG	627	
Db	271598	GGCGAGAGCCCGCAGAAGTGSACCCAGCTGGCGTCCGAGATCCGGGTGTCAACCGAGAG	271539	
Qy	628	ACAACAACAGGTGTCCACCGCTCTACACAGCTCGAGAAAGAGGGCAAACTCCTCTTCCCA	687	
Db	271538	ACCAAGACCGGGTTCACCGCTGTACGAGTGCACGCGAGCGACACCTGCTTCCCG	271479	
Qy	688	GCCATCAACGTCAACGACGCTGTTCACAAAGTCCAAAGTTCGATAACATCTACGGTGC	747	
Db	271478	GGCATCAAGCTCAACGACGCGTCAACAAAGTTCGACAAAGTTCGACAAAGTTCGCGT	271419	
Qy	748	CACCTCCCTTATCGATGGTATCAACCGTCTTCGATGTCATGATCGCGGCAAGACAGCT	807	
Db	271418	CACCTCCGTATCGACGGCATCAACCGCGCACACGAGTCTCTGATCGGGGCAAGACG	271359	
Qy	808	CTCGTCACTGGGTTACGGCGATTCGGCAAGGGTCTGGCTCAATCCCTCCGTGGCAAGGC	867	
Db	271358	GTCTGTCTGGGCTACGGCGACGCTCGCAAGGGTCTGGCGAGTCCCTGGCGGCCAGGC	271299	
Qy	868	GCTCGGTTATATCAACAACTCGACCCATCTGGCTCTCGAGCTGCCATCGAAGGC	927	
Db	271298	GGCGCGGTATCATACCCAGATCGACCCGATCTGGCGCTCTGCAGCGCGCATCGACGC	271239	
Qy	928	TACCAAGTCCCGCGCATCGAGAAAGTCTCAAGGATCTCGATCTTCTGTTACATGCACA	987	
Db	271238	TTCCAGGTCCACCGCTGACGAGTCTGCAACAGCCGACATCTTGTCAACACGACC	271179	
Qy	988	GGAACTCGGATATCATCTCTTTGACATGATGCGCCGACATGAAGATAGAGCTATTGTC	1047	
Db	271178	GGCAACAGGACATCATATGCGCAAGACATGCGCAAGATGAAGCAACGAGCATCGTC	271119	
Qy	1048	GGTAACATCGGCACATTGCATAAAGAAATGATACAGATGGCTCATGAATACCCAGGC	1107	
Db	271118	GGCAACATCGGCCATTTCGAAACAGATCGACATGGCGGGCTCGCCAGACCCCGGC	271059	
Qy	1108	ATCAAGCACATCCCAATCAAGCCAGAATAACGACATGTGGGAATCCAGATGGCCACGCT	1167	
Db	271058	ATGCTCAAGGACGAGGTCAAGCGCAGGTCCACACTTGACCTACCCGCGCAAGGTG	270999	
Qy	1168	ATCTCTCTTCTGTCGAGGCGCTCTTAACTTGCTGGCTGGCTACAGGTACCCATCT	1227	
Db	270998	CTCATCGTCTCTCGAGGGCGCTCTGTGAACCTGGGCAAGCCACCGGCCACCCGCTG	270939	
Qy	1228	TTCTGTATGTCAATGTCTATCAAAACACAGACATCGCTCAGCTCGACTCTACGAAAG	1287	
Db	270938	TTCTGTATGTCCAACTCTTTCGGGACACAGACCTGGCCAGATCGAGCTGTACCAAG	270879	
Qy	1288	AGAGGAATCTCGAGAAGAGTTTACACATTCCGAGCGATCTCGATGAAGATGTGCT	1347	
Db	270878	CCCGACGAGTACCGGACCGACGCTTCACTGTGTGCCCAAGCACCTGGAGGAAAGTGC	270819	
Qy	1348	CGCTTCCACCTCGGATCTCTCGATGTCCACTTCAAAAGCTTACACAGAACGAGCTGAC	1407	
Db	270818	CGGCTCCACTCTCACTCCCTCGGGTCAAGCTGACACGCTCGCGCGGACGCGCCAC	270759	
Qy	1408	TACATCAACGTTCCAGTTGAGGGTCTTTACAAAGTGTGATGCTTACCGTTATTAA	1461	
Db	270758	TACTTCGGGTCAAGTTCGAGGGCCCTTACAAAGCGGACCATCGGCTACTGA	270705	

RESULT 4  
AP005041  
LOCUS 300425 bp DNA linear BCT 10-MAY-2003  
DEFINITION Streptomyces avermitilis genomic DNA, complete genome, section 21/30.

```

ACCESSION      AP005041 BA000030
VERSION        AP005041.1 GI:29608607
KEYWORDS
SOURCE
ORGANISM
Streptomyces avermitilis MA-4680
Streptomyces avermitilis MA-4680
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
1
AUTHORS
Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Takahashi, C.,
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,
Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
TITLE
Genomic sequence of an industrial microorganism Streptomyces
avermitilis: deducing the ability of producing secondary
metabolites
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
21477403
JOURNAL
MEDLINE
PUBMED
11572948
REFERENCE
2
AUTHORS
Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H.,
Shiba, T., Sakaki, Y., Hattori, M. and Omura, S.
TITLE
Complete genome sequence and comparative analysis of the industrial
microorganism Streptomyces avermitilis
Nat. Biotechnol. 21 (5), 526-531 (2003)
22608306
JOURNAL
MEDLINE
PUBMED
12692562
REFERENCE
3
AUTHORS
Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,
Kushida, N., Director-General of Biotechnology Center, Shiba, T.,
Sakaki, Y. and Hattori, M.
TITLE
Direct Submission
Submitted (29-MAR-2002) Director-General of Biotechnology Center,
National Institute of Technology and Evaluation, Biotechnology
Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
(E-mail: bioente.go.jp, URL: http://www.bio.nite.go.jp/,
Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)
This work was done in collaboration with Haruo Ikeda (*1), Jun
Ishikawa (*2), Akiharu Hanamoto (*3), Chigusa Takahashi (*3), Mayumi
Shinose (*3), Hiroshi Horikawa (*4), Hidekazu Nakazawa (*4), Tomomi
Osonoe (*4), Norihiro Kushida (*4), Hisashi Kikuchi (*4), Tadayoshi
Shiba (*5), Yoshiyuki Sakaki (*6, *7), Masahira Hattori (*1, *7)
and Satoshi Omura (*1, *3).
Final finishing process and all annotation were done by H. Ikeda
and J. Ishikawa.
*1 Kitasato Institute for Life Sciences, Kitasato University
*2 National Institute of Infectious Diseases
*3 The Kitasato Institute
*4 National Institute of Technology and Evaluation
*5 School of Science, Kitasato University
*6 Institute of Medical Science, University of Tokyo
*7 RIKEN, Genomic Sciences Center
Following url is also available.
http://avermitilis.is.kitasato-u.ac.jp.
Location/Qualifiers
1. .300425
/organism="Streptomyces avermitilis MA-4680"
/mol type="genomic DNA"
/strain="MA-4680"
/db xref="taxon:227882"
/notes="This strain is also named as strain: ATCC 31267,
NCIMB 12804 or NRRL 8165.-synonym: Streptomyces
avermectiniius"
162. .383
/genes="infA3"
162. .383
/genes="infA3"
/notes="SAV4949"
/codon start=1
/transl table=11
/product="putative protein chain initiation factor IF-1"
/protein_id="BAC72661.1"
/db xref="GI:29608608"
/translation="NAKKQGAIEIEGTWVPSLPMFKVLQNGHQVLAHISGKMRH"
gene
CDS
YIRILPDDRVVVELSPYDLTRGRIVYRK"
443. .556
/genes="rpmJ"
443. .556
/genes="rpmJ"
/notes="SAV4950"
/codon start=1
/transl table=11
/product="putative ribosomal protein L36"
/protein_id="BAC72662.1"
/db xref="GI:29608609"
/translation="HKVKPSVKIKCDKCRVIRRHGRVVICDNPRIHQKQG"
760. .1140
/genes="rpsM"
760. .1140
/genes="rpsM"
/notes="SAV4951"
/codon start=1
/transl table=11
/product="putative ribosomal protein S13"
/protein_id="BAC72663.1"
/db xref="GI:29608610"
/translation="WAVSGVDIPREKRVVEALTYVFGIGRTLSOKLAEQVNPNT
VRDLSRELVRIRREVNNKTEGDLRREVQADIRKVEICCYQGLRHRGLFVHGQR
TSTNARTKGRRAIACKKFGKK"
1238. .1642
/genes="rpsK"
1238. .1642
/genes="rpsK"
/notes="SAV4952"
/codon start=1
/transl table=11
/product="putative ribosomal protein S11"
/protein_id="BAC72664.1"
/db xref="GI:29608611"
/translation="MPPKRGQGAACKVRKKEKNVAHGHAIKSTFNNTIVSITDPSG
NVTWSAGHVGFSGSRKSTPFAAQWAAFSARRAQEHGKVDVVFVKFGSGRETAI
RSLQATGLVSGISQDVTFTTHNSCRPPKRRV"
1785. .2807
/genes="rpoA1"
1785. .2807
/genes="rpoA1"
/notes="SAV4953"
/codon start=1
/transl table=11
/product="RNA polymerase alpha subunit"
/protein_id="BAC72665.1"
/db xref="GI:29608612"
/translation="MLIAQRPSLTERVDFRFRFVLEPFGVTLGNSLRRTLLS
SIPGAAVTSIRIDGVLFHEFTVPGEKEDVTDLLINKLVVSSEHDEPVMYLRKQGP
GLVTADIAPGAVEVENVPLVLAUNGKLEMLIVRGRGVSAVONKQVGEIG
RIPVDSIVPKVTVKVEATRVQRTDFDKLIVDVTQKMRPRDAVONKQVGEIV
FGLARELNDIAEGIDMGFSPDAAALDPIELELTVRSYNCLKREGIHSVGLV
ARSEALDLIRNFCAKSIDEVKAKLAGMGLAKDSPFPDPTAAADAFGADDDADG
VETEYQ"
2989. .3492
/genes="rplQ"
2989. .3492
/genes="rplQ"
/notes="SAV4954"
/codon start=1
/transl table=11
/product="putative ribosomal protein L17"
/protein_id="BAC72666.1"
/db xref="GI:29608613"
/translation="WPKPTKGARLGSAHAKLLIANLAKSLFEGHRIITTTAKARRL
RPAERLVTKAKGDLHNRQVLOVITDKSVVHTLFTIEIGRVENRPGGYTRITKIG
RQGDNPMAVTELVLEALTVAAQTGAEATKRSAKDAEAKTEAPREVVEDTKPADEA
ARESKDA"
3582. .4439
/genes="trnA"
3582. .4439
/genes="trnA"

```

```
/note="SAV4955"
/codon_start=1
/transl_table=11
/product="putative pseudouridine synthase"
/protein_id="BAC72667.1"
/db_xref="GI:29608614"
/translation="MSDEVPQGFVRVRLDLSYDGFSSGAWQAQAGRTVQGEIEDAL
RTVRSGETYEUTVAGRTDAGVHARVDLPGLMAEEQEKLLKRLAGLRPKDV
RWYSLTAPSGFNARSAIRRYAVRQVNTGSDVPLASHVLMHWDPLDYDMNBA
RLVEGHDPAAYCKKEGATTTIRLOELSLVRGDDGVITATVADAPCHNWRSLIGA
LLFVGDGHRGSDWPKGLAVGVRSVAVHVRPHGLTLEVGYPADPELLAAENKRNK
RLPAGCC"
complement(4444..5331)
/note="SAV4956"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC72668.1"
/db_xref="GI:29608615"
/translation="MRPMSYGGGEPFDPKPSQGFQWGNQTDQTPDMAALAESES
RNKRRLLIIVGAFATVAIGAAVAVVVSANDTEASKNPANLPATDIPSSSTAPA
PFDFTTAPPLPKDPIISAKKDAPLSPDILFPSTELTIGRAYOKGATPSTNCA
SAVQGTLPQALTNSGTRVVRVYSGKGVAVTVGAVPDTAKAKAKTADNKSII
TSLPQGVAAFCRTAVCRSTTNAYGRVAYFTLAGFTNGDKVTEKDAAVFTTGDDLABY
TPRQIRQGEARQASANK"
5704..6147
/gene="rplM"
CDS
5704..6147
/gene="rplM"
/note="SAV4957"
/codon_start=1
/transl_table=11
/product="putative ribosomal protein L13"
/protein_id="BAC72669.1"
/db_xref="GI:29608616"
/translation="MTYSPKPGDITROMYVIDQVVLRLATTAANILRKHKPIY
APHVDAGFVIIINADQVHLNGKTKTOKLAVHSYVPGGLSRVRYDELLAKNPEAVE
KALKQWIPKNTLRQVLSKLKLVYSGDQHPHAAQPPVPEITQVAQ"
6190..6711
/gene="rpsI"
CDS
6190..6711
/gene="rpsI"
/note="SAV4958"
/longer_N-terminus_compared_with_other_bacteria"
/codon_start=1
/transl_table=11
/product="putative ribosomal protein S9"
/protein_id="BAC72670.1"
/db_xref="GI:29608617"
/translation="MAETTVQPVVEETVVDVIEYTTSESPVPEGEYTSVSRFGE
POPAAGLRKNALVRVIVPGTKWKGRTLEDYFPNKHVQQEVNEFPKVELEGR
Query Match 44.0%; Score 642.4; DB 1; Length 300425;
Best Local Similarity 66.0%; Pred. No. 1.2e-139;
Matches 947; Conservative 0; Mismatches 481; Indels 6; Gaps 1;
QY 34 GAGTACAGATTCGCGACATCACTCCATGTTCTGCGCGTAGGAACTTACCTTGT 93
Db 119232 GACTTCAGGTGCGCGATCTCTCCGCGCGAGTTCGCGCGCAGAGATCACCCTCGCC 119291
QY 94 GAGAAGAAATCCAGGCTTTATGGTTCTTCGTGAGCGTTATTCCGCTTTAAGCCATTG 153
Db 119292 GAGCAGAGATCCCGGCTGATGTCGATCCGAGGAGTACGCCAGGCGGAGCCCTC 119351
QY 154 AAGGCTGTCAGATCTGTTGTTCCCTCCATGACAGTCCAGACAGCGCTCTCATCGAG 213
Db 119352 GCGGCGCGCCGCGTCAACCGCTCCCTGACATGACGGTGCAGACCGCGTACTCATCGAG 119411
QY 214 ACATCAGAGCTCTTGGTGTGATGTCAGATGGGCTTCTGCAACATCTTCTACAAA 273
Db 119412 ACCCTGGTGGCTGGCGCGGAGTCCGCTGGGCTCTCTGCAACATCTTCTCCACCCAG 119471
QY 274 GATACAGCGCGTGTGCTATCGTTGTGCGGCCAACAGGACACACAGAGAGCGCGGT 333
```

```
Db 119472 GACCAACGCGCGCGCCATCGCGTGGGCCGGAACGGCAACGCCCGCAGCAACCCCGCAGGC 119531
QY 334 ATCCACAGTCTTCCCTGGAAGGGGGAACACTCCCAAGATACTGGGGAACACATACCCG 393
Db 119532 GTCCCGGTTCTTCCCTTGGAAAGGGCGAGCCTGGAGGAGTACTGGTGTGACGGAGCAG 119591
QY 394 GCTCTTCATGGCCAGAGGTCAA-----GGCCCCACAGCAGTTTGTCTGATGATGTTG 447
Db 119592 GCGCTGACCTGGCCGAACACCCCAACCGGGGCGCCGAAATGATCTCTGACAGCGCGGT 119651
QY 448 GATGCTACATCTCTCATCTCTCAAGGCTTCGAATTCGAACACAGCGGTGCTGTTCCAGAG 507
Db 119652 GAGCGCAACCTCTCTCTGTCACAAAGGGGTGAGTACGAGAAGGCGCAAGTTCCGTCC 119711
QY 508 CCAACAGAGCTGACAACTCGAATACCGTGTGGTTCTTGTACTACACTCAAGCAGGTCTTC 567
Db 119712 GTTCGACACCGGGAGTCCGACGAGCACCGCGTCTATCTCGAACTCTTCGACCGCACATC 119771
QY 568 AACCAAGACAGAACCACTGGCACACAGTTGCTGCGCGCATGAACGGTGTTCGGAAGAG 627
Db 119772 ACCGACGCTCGAGAAAGTGACCCAGCAGCTCGCTCGAGATCCGCGCGTACCGAGAG 119831
QY 628 ACACACACAGGTCTCCACCGCTCTTACAGCTCGAGAGGAGCAAACTCTCTCTCCCA 687
Db 119832 ACCAGCGGGGTGACCCGCTGTACGAGTGCACGGCAGCGGACCCCTCTCTGTTCCG 119891
QY 688 GCATCAACGTCAACGAGCTGTTCACAAAGTCCAAAGTTCGATAACATCTACGGTGCCTCC 747
Db 119892 GCGATCAACGTGAACGAGCGCGCTACCAAGTTCGAAGTTCGACAAAGTTCGGGTGCGC 119951
QY 748 GACTCCCTTATGATGATGATCAACCGTCTTCGAGTTCATGATCGGCGGCAACAGACT 807
Db 119952 CACTCCCTGATCGACGGCATCAACCGCGCACCGAGCTCTGATCGGCGGCAACCGCG 120011
QY 808 CTGCTCATGGTTACGGCATCTCGGCAAGGGTGGCTCAATCCCTCCGTGGGCCAAGGC 867
Db 120012 GTCTCTTGGGCTATGGCGAGCTGGCAAGGCTGTGGGAGTCTCTCGCGCGCAGGC 120071
QY 868 GTCGCGTTATCTACAGAACTCGACCCCAATGTCGCTCTCCAGGTGTCATGGAAGGC 927
Db 120072 GCGCGGTGATCGTCCAGAGATCGACCGATCTGCGGCTCGAGGCGCGATGCAAGGC 120131
QY 928 TACAGGTCCGCGCATCGAGAGTCTGCAAGATGTCGATCTTCTGTTACATGACACA 987
Db 120132 TACAGGTACAGACCTTCGACGAGTCTGCAAGGCTGATGCAAGGCTGATCTTCTGTCACAGAG 120191
QY 988 GGAACTCGGATATCTCTCTGTTGACATGATGGCCCGCAGATGAAGATAAGCTATTGTC 1047
Db 120192 GGCACAAAGGACATCATCATGCTCGGACATGSCCAAGATGAAGCACAGGCGATCGTG 120251
QY 1048 GGTACATCGCGCACTTCGATTAACGAATTTGATACAGATGGCTCATGAATACCCAGGC 1107
Db 120252 GGCACATCGCGCACTTCGATTAACGAATTTGATACAGATGGCTCATGAATACCCAGGC 120311
QY 1108 ATCAAGACATCCCAATCAAGCCAGATACGACATGTGGGAATTCCTCAGATGGCCAGCT 1167
Db 120312 ATGCTCAAGACGAGTCAAGCCGAGGTCCACACTTGAAGTTCCTCCAGCGCAAGTTC 120371
QY 1168 ATCTCTCTTGTGTGAGGCGCGCTTCTTAACCTTGGCTGCTGAGTTCAGCTACCCATCT 1227
Db 120372 CTCTCTCTTGTGTGAGGCGCGCTTCTTAACCTTGGCTGCTGAGTTCAGCTACCCATCT 120431
QY 1228 TTCGTTATGTCATGTCATTAACCAACAGACACTGCTCAGCTCAGCTCAGCTCAGTAAAG 1287
Db 120432 TTCGTTATGTCATGTCATTAACCAACAGACACTGCTCAGCTCAGCTCAGCTCAGTAAAG 120491
QY 1288 AGAGAAATTCGAGAAAGAGTTTACACTTCCGAAGCATCTCGATGAAGAGTCTGCT 1347
Db 120492 CCCGACGAGTACCCGACCGAGCTCTACGTTGTCGCCAAGCACTCGACGAGAGTCTGCC 120551
QY 1348 GCGCTCAGCTCGGATCTCTGATGTCAACCTTACAAAGCTTACAGAGAGAGGCTGAC 1407
```

```
Db 120552 CGCTCCACTCGACTCGCTCGGCGTGAAGCTGACGAGCTCGCCGCCGAGCAGCGCCGCG 120611
Qy 1408 TACATCAGAGTTTCAGTTGAGGCTCCTTACAAAGTCTGATGCTTACCGTTATTA 1461
Db 120612 TACATCGGCGTGCAGGTCGAGGCGCGCTTCAAGTCGGACACACTACCGCTACTGA 120665

RESULT 5
SAR416377 3800 bp DNA linear BCT 15-OCT-2001
LOCUS Streptomyces argillaceus mtmZ gene, mtmA gene and mtmH gene.
DEFINITION Streptomyces argillaceus mtmZ gene, mtmA gene and mtmH gene.
ACCESSION AJ416377
VERSION AJ416377.1 GI:16197721
KEYWORDS adohcyase; AdoMet synthetase; MTHF reductase; mtmA gene; mtmH gene;
SOURCE mtmZ gene; thioesterase.
ORGANISM Streptomyces argillaceus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE
1 Fernandez Lozano, M., Lombo, F., Brana, A. F., Salas, J. A. and Mendez, C.
AUTHORS Generation of a Streptomyces argillaceus mutant by deleting two
TITLE methyl cycle-related genes located within the mithramycin gene
JOURNAL cluster leads to higher production of mithramycin
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 3800)
JOURNAL Mendez, C.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-2001) Mendez C., Biologia Funcional e IUOPA,
JOURNAL Universidad de Oviedo, C/ Julian Claveria s/n, 33006 Oviedo, SPAIN
FEATURES
source
1..3800
/organism="Streptomyces argillaceus"
/mol_type="genomic DNA"
/db_xref="taxon:41951"
7..789
/gene="mtmZ"
7..789
/gene="mtmZ"
/function="putative role in polyketide chain biosynthesis
and release"
/codon_start=1
/transl_table=11
/product="putative thioesterase"
/protein_id="CAC94888.1"
/db_xref="GI:16197722"
/translation="MAAQPIRLYCFAHAGAGVSFHWAAVGPDPVVPVLLPGRD
RRREPRVTRALLDDVMRLFGGAPTAPFVLYGHSLGALVAYTVGRALHEAGRGFA
LLAVGACPPPLPPEPAYALGVDPDAELRLTDLMAVPSGTGTGAGGIWDRAVRPVL
RDULTARALRTAACTSSPADPLTTPELLIYAADDPMASPAVAGWSRWSGPRRLHT
VSGGHFFVRGPDELPRLLGQACALPARRTPRAAGPVTSTGN"
918..2300
/gene="mtmA"
918..2300
/gene="mtmH"
/function="putative role in an activated methyl cycle"
/codon_start=1
/transl_table=11
/product="AdoMet synthetase and MTHF reductase fusion"
/protein_id="CAC94889.1"
/db_xref="GI:16197723"
/translation="MSGHLFTSESVEGHPKIAIRISGTPIDALLREDPRSRVAVET
LITTCQVAVADKVTGAVVPIAOLVRDAVLITGYDSSAKFGDGCAGCVSVIGPQSD
IAQGVDFAVSLVEGAAAGADIDELDRQADGQGLMPCGACNDESINLPLIELNRL
ARLLIDVRNGLPPTFARTARPGSPEDRAGARVIMDTIRRGEARLPDFVSVTYEAGG
SPRDTVALTRIAAETTLRVAHLTAVGHSVAELRQITGRVADYGRIDRVFALRGDP
GDGPGSPHFEGTFVAHELVEFSLGDFTHIRMAAPFPHRPSRPMWSDIRHFWKVC
RAGADYATOMFGPGCYLRLRDPAAAGCHIPIIPEIMPAMDVQIRRFALISAVAF
TDLAHLRLEQARGNPVAGHRIGVHDHATMAERLLABGVPLCHITLNRSTALDIHN
LTSRSLSVR"
2290..3738
/gene="mtmH"
2290..3738
```

```
/gene="mtmH"
/function="putative role in an activated methyl cycle"
/codon_start=1
/transl_table=11
/product="adohcyase"
/protein_id="CAC94890.1"
/db_xref="GI:16197724"

BASE COUNT 590 a 1485 c 1197 g 528 t
ORIGIN

Query Match 42.1%; Score 615.2; DB 1; Length 3800;
Best Local Similarity 65.3%; Pred. No. 2.5e-133;
Matches 937; Conservative 0; Mismatches 488; Indels 2;

Qy 34 GAGTACAGAAATTCGGACATCAACCTCCATGTTCTCGCGCTAAGGAACTTACCTTGCT 93
Db 2308 GACTTCAAGTTCGCCGATCTCTCCCTGGCCGAGTTTCGCCGCAAGAGATCACCTCGCC 2367

Qy 94 GAGAAGGAATGCCAGGTCTTATGTTCTTCGTGAGCGTTATTTCGCTTCTAAGCCATTG 153
Db 2368 GAGCAGGATGTCGCGCTGATGTCGATCGCAGGAGTAGCCGAGGCCACGCCCTC 2427

Qy 154 AAGGTTGTCAGATCTCTGTTCCCTCCATGACAGTCCAGACAGCGTCTCTCATGAG 213
Db 2428 GCGCGCGCGCGCTCAACCGGCTCTCTGCATGACAGTCCAGCCGCTCTCTCATGAG 2487

Qy 214 ACATCTCAGAGCTTGTGTGTCGATGTCAGATGGCTTCTGTCAACATCTTCTTACACA 273
Db 2488 ACCCTGTGCGCTCGCGCGCCAGGTCCGTGGGCTCTCTGCAATCTTCTTACCCAG 2547

Qy 274 GATACAGCGCTGCTGCTATGCTTGTGCGCCCAACAGGCAACAGGAGACCCAGCGGT 333
Db 2548 GCGCCAGCGCGCGCGCTATGCTGCGCCCGAGCGCACGCCGACACACCCGCGGCC 2607

Qy 334 ATCCAGTCTTCCGCTGGAGGCGCAACACTCCAGATACTGGGAGAACACATACCGC 393
Db 2608 GTCCCGGTCTTCCGCTGGAGGCGCGAGCGCTGAGGAGTACTGTGTGTCACCGAGCAG 2667

Qy 394 GCTCTCATGCGGCAGATGGTC-----AAGGCCCAACAGCAGGTTGTGATGATGTTGTT 447
Db 2668 GCGCTGACCTGGCGGAGCAGCCGCCACCGCGGCCCGGAGCATGATCTCTGACGAGCGGT 2727

Qy 448 GATGCTACATCTCTATCTTCCAGGGCTTCGAATTGAAACAGCGGTGCTGTTCAGAG 507
Db 2728 GACGCCACTTGTGTTGCCAAGGGCGTTCGAGTACGAGAGCGGCAAGGTCCTCCCTCG 2787

Qy 508 CCAACAGAGTGCACCACTCGAATACCGTCTGCTTCTTACACTCAAGCAGGTCTTC 567
Db 2788 GTGACACCGCGGAGAACGACGAGCAGCCGCTGATCTCTGAACTCTCAACCGCACCATC 2847

Qy 568 AACCAAGAACAAGAACCACTGGCACACAGTTGCTGCGGCATGAACGGTGTTCGGAAGAG 627
Db 2848 ACCGACGCGCTCCCAAGTGGAGCGCAGCTGGCGTCCGAGATCCGCGCGGTGACCGAGAG 2907

Qy 628 ACAACACAGGTGTCCACCGCTCTACACGCTCGAGAGGAGGGCAAACTCTCTTCCCA 687
Db 2908 ACCACACCGCGCTCCACCGCTCTGACAGATGACGCGGAGCGGTGCTCTCTTCCCC 2967

Qy 688 GCATCAAGTCAACAGCGCTGTTCRAAGTCCAAAGTTCGATTAACATCTACGGTGGCGC 747
Db 2968 GCGATCAAGTCAACAGCGCGGTTCACCAAGTCCGAAGTTCGAACATCAAGTACGGTGGCGC 3027

Qy 748 CACTCCCTTATCGATGTTATCAACCGTCTTCCGATGTCTATGATCGGCGGCAACAGCT 807
```



```
Db 3028 CACTCGCTGATCGACGGCATCAACCGCGCCACCGACGCTCCTGATCGGGCGCAAGACGGCC 3087
QY 808 CTCGTATGGTTACGGGATGTGGCAAGGGCTGGCTCAATCCTCTCGGTGGCCCAAGGC 867
Db 3088 GTGGTCTTTCGGCTACGGGACGCTGGCAAGG---CTGCGCGAGTCTCGCGGGCCAGGGC 3144
QY 868 GCTCGCGTATCATACAGAACTCGACCCATCTGGCTCTCCAGGTCGCATGGAAGC 927
Db 3145 GCCCGGCTCATCTGACCCAGATCGACCGGATCTGGCACTCGACGGCGGCGATGAGCGG 3204
QY 928 TACCAGTTCGCGCGCATCGAGGAAGTCTCAAGGATGTCGATATCTTCGTATCATGCACA 987
Db 3205 TACCAGTTCAGGACCTTCGACGAGTCTGACAGAGCGGACATCTTCATCACACGACC 3264
QY 988 GGAACCTGCAATCATCTCTGTCACATGATGGCCGACATGAAGATAGGCTATTGTC 1047
Db 3265 GGCACAAAGGACATCATATGCGCGCGGACATGCGCAGGATGAAGCACCGGCGCATCGTC 3324
QY 1048 GGTAAACATGGCCCACTTGATAAGAAATTGATACAGATGGCTCATGAATATCCCAAGC 1107
Db 3325 GGGAAACATCGGCCACTTCGACAGAGATCGACATGGCGGCTCGCCAGGTCCCGCGC 3384
QY 1108 ATCAAGCACATCCCAATCAAGCCAGAAATGATACAGATGGGAAATCCCAAGATGGCCAGCT 1167
Db 3385 ATCGTCAAGGACGAGATCAAGCGCAGCTCCACACCTGGACCTCCCGGACGGCAAGAG 3444
QY 1168 ATCTCTCTTCTGCTGAGGGCGCGCTTTAACTTGGCTGGCTGCTACAGGTACCCATCT 1227
Db 3445 ATCATCTGTGTGTCGAGGGCGCTTCTGCTGAACCTGGGCAACCGCCACCGGCGCCACCGCTCG 3504
QY 1228 TTGCTTATGTCATGTCATTACAAACCAAGACACTCGCTCAGCTCGACCTCTACGAAAG 1287
Db 3505 TTGCTGATGTCGAACTCTTCGGGACCAAGACCTCGCCACATCGAGCTTTCACCAAG 3564
QY 1288 AGAGAAATCTGAGAAAGGTTTACACTTCCGAAACATCTCGATGAGAAAGTCGCT 1347
Db 3565 CCCGCGCACTACCGGACCGGCTTACACCTGCGCAAGCACTCGACGAGGTCGCGC 3624
QY 1348 CGCTCCACCTCGATCTCTGATGTCACCTTACAACTTACACAGAGCAGGCTGAC 1407
Db 3625 CGCTCCACCTGACGCGCTCGGGTGAGCTCACGACCTCCGCGCGGACGAGGCTCG 3684
QY 1408 TACATCAACCTTCAGGTCCTTACAGTCTGATGCTTACCGTTATTAA 1461
Db 3685 TACTCGCGTTCAGCTCGACGCGCGTACAGTTCGACCACTACCGTACTGA 3738

RESULT 6
AX063937
LOCUS AX063937 1557 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 219 from Patent WO0100843.
ACCESSION AX063937
VERSION AX063937.1 GI:12541649
KEYWORDS
SOURCE Corynebacterium glutamicum
ORGANISM Corynebacterium glutamicum
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Corynebacteriaceae; Corynebacterium.
REFERENCE 1
AUTHORS Pompeius M., Kroeger B., Schroeder H., Zelder O. and Haberhauser G.
TITLE corynebacterium glutamicum genes encoding metabolic pathway
JOURNAL Patent: WO 0100843-A 219 04-JAN-2001;
FEATURES
source BASF AKTENGESSELLSCHAFT (DE)
location/Qualifiers
locus_tag AK1557
organism="Corynebacterium glutamicum"
mol_type="genomic DNA"
db_xref="taxon:1718"
101..1537
/notes="RXN00132"
/codon_start=1
/transl_table=11
CDS
```

```
/protein_id="CAC25210.1"
/db_xref="GI:12541650"
/translation="MAQWDFKVADLSLAEGHQIRLAIEYEMPLQMLRKEFADEOP
LKGRIAGISIMTVOTAVLIETLTALGABRWASCNIFSTQDEAAAIIVVSGTVEEP
AGVPVPAKGSLEBEVWCINQIFSGWDELNMLDDGGBATMAVIRGREYQAGLVP
PARANDSDEYIAFLCMLREVLAEPGKWKAEAVKGVTEETTTGVRHRIHFAESGVL
FPFAMNDVATKSKFDNKGTHSLIDGINRATMLMGGKNVLVCGYGDVGKGAEEA
FDQGARVKVTEADINALQALMDYSVVTVDEAIEDADIVITATNKDIISEQMLK
MDHALLGNIHFDEIDMHSLLHRDVRTTIRKPOVDEFTFTSGRSIIVLSEGLLN
LGNATHPSFVMSNSPADITIAQIELFQNEGYENEVRLPKVLDEKVAIRIHEVALGG
QLUTELTKQAEYIGVDVAGFPKPEHRY"
BASE COUNT 348 a 440 c 431 g 338 t
ORIGIN
Query Match 40.8%; Score 596.4; DB 6; Length 1557;
Best Local Similarity 64.5%; Pred. No. 6.1e-129;
Matches 923; Conservative 0; Mismatches 501; Indels 6; Gaps 2;
QY 32 TCGAGTACAGAAATTCGCCACATCAACCTCATGTTCTGCGCGTAAGAACTTACCCCTTG 91
Db 114 TGGACTTCAAGGTTCCCGATCTTTCACTAGCAGAGCGACGCTCACAGATTCTGCTTGG 173
QY 92 CTGAGAAGGAAATTCGACGCTCTTATGTTCTTCTGAGCGTTATTTCGCTTCTTAAGCCAT 151
Db 174 CAGAGTATGAGATGCGAGGCTCTCATGAGTTCGCAAGGAATTCGACAGACGACGCTT 233
QY 152 TGAAGGTTGTGAGAAATCTCTGTTTCCCTCCATGACATGACAGTCCAGACAGCCGCTCTCATCG 211
Db 234 TGAAGGCGCCCGAAATTCCTGTTCTTATCCATGACGCTCCAGACCGCGCTGCTTATTG 293
QY 212 AGACACTCACAGCTCTTCTGCTGATGTCAGATGGGCTTCTCGAAATCTTCTCTTACAC 271
Db 294 AGACCTCTACTCTCTTGGGCGCTGAGGTTCTGTTGGGCTTCTTGCACATTTCTTCCACCC 353
QY 272 AAGTATACAGCCGCTCTCTATCTGTTCTGCGCCCAACAGCGCACACAGAGAGCCAGCGC 331
Db 354 AGGATGAGGCTCGACGGGCTATCGTTGTGGCG---TCCGGCACCGTCGAGAGCCAGCTG 410
QY 332 GTATCCGAGTCTTCCGCTTGAAGGGCGAAACACTCCCGAGATATCTGGAGAAACACATACC 391
Db 411 GTGTTCCAGTATTCGCGTGGAAAGGTTGAGTCACTGAGGAGTACTGTTGGTGTGATCAACC 470
QY 392 GCGCTCTCATATGCGCAGATGTTCAAGGGCCACAGCAGGTTTGTGATGATGGTGTGATG 451
Db 471 AGATCTTTCAGCTTGGGGCGATGAGC---TGCCAAACATGATCTCGACGAGCGGGTGAAG 527
QY 452 CTACACTCTCTCATCTCCAAGGGCTTCGAATTCGAAACAGCGCGCTGCTGTTCAGAGACCAA 511
Db 528 CCACCATGCTGTATTTCGGGTCGCGAATACGAGCAGGCTGGTCTGTTCCACACGACGAG 587
QY 512 CAGAGCTGACAACTCGAATACCGCTGGCTTCTTGTCTACTACTCAAGCAGGTTTCAACC 571
Db 588 AGGCCCAACGATTCGATGAGTACATCGCATTTCTTGGGATGCTCGGTGAGGTTCTTGTG 647
QY 572 AAGAAGAAGAACCACTGGCACACAGTTCGTCGGCATCAACGGTGTTCGGAAGAGACAA 631
Db 648 CAGAGCTGCGAAGTGGGGCAGATCGCTGAGCGCTTAAAGGTTGACCCGAGAAACCA 707
QY 632 CAAACAGTGTCCACCGCTCTTACAGCTCGAAGAGGAGGCAAACTCTCTTCCCAAGCA 691
Db 708 CCACCGGTGTGACCGCTGTACCACTTCGCTGAAGAAGGCGTCTGCTTTCACAGGA 767
QY 692 TCAAGCTGACACGCGCTGTTACAAAGTCCAAAGTTCGATACATCTACGCTGCGCGCTACT 751
Db 768 TGAACGCTCAACGACGCTGTCAACAGTTCCAAGTTTGATACAGTACGCGCACCGCCACT 827
QY 752 CCCTTATCATGCTATCAACCGTGTCTTCCGATGTCATGATGGGGCAAGACAGCTCTCG 811
Db 828 CCCTGATCGACGCTCATCAACCGCGCTCTGACATGCTCATGGGGCGGCAAGCTGCTTG 887
QY 812 TCATGGGTTACGGCGATGTGCGCAAGGGCTGGGCTCAATCCCTCCGTGGCGCAAGGCGCTC 871
Db 888 TCTGCGGTTACGGCGATGTGCGCAAGGGCTGGGCTTTCGAGGCTTTCGACGCGCGAGGCGCTC 947
```

```
QY 872 GCGTTATCATCAAGAACTGACCAATCTGCGCTCTCCAGGTGCGCATGAGGCTACC 931
Db 948 GCGTCAAGGTACACCGAAGCTGACCAATCAACGCTCTTCAGGCTCTGATGGATGCTACT 1007
QY 932 AGTCTCCGCGATCGAGAACTGCTCAAGGATGTGATATCTTCGTTACATGACAGGAA 991
Db 1008 CTGTGTACCGTTGATGAGGCAATCGAGGACCGCATCGTATGATACCGGACCGGCA 1067
QY 992 ACTGCGATATCATCTCTGTTGACATGATGCGCCAGATGAAGGATTAAGGCTATTGTGCGTA 1051
Db 1068 ACAAGGACATCATTTCTTCGAGCAGATGCTCAAGATGAAGATCAAGCTCTGCTGGCA 1127
QY 1052 ACATCGGCACTTCGATACGAAATGATACAGATGCGCTCATGAATACCCAGGATCA 1111
Db 1128 ACATCGGTCACTTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1187
QY 1112 AGCACATCCCAATCAAGCCAGATACGATGTTGGGAATCCAGATGCGCAGCTATCC 1171
Db 1188 CCGGACACGATCAAGCCACAGGTGACAGGTTCACTTCTCCAGGTGCGTCCATCA 1247
QY 1172 TCCTTTCTGAGGCGGCTTTTAACTTTGGCTGGCTTACAGGTCAAGGTCACCATCTTTTCG 1231
Db 1248 TCGTCTGTCGAGGCTGCTGTTGAACCTTTGGCAAGCCACCGGACACCATCATTTG 1307
QY 1232 TTATGTCATGTCATCAAAACAGACACTGCTCAGCTCAGCTCAGCTTACGAAAGAGAG 1291
Db 1308 TCATGTCCAACTCTTTGCGCGATCAGACCAATTCGCGAGATCGAATCTGTTCCAAACGAAG 1367
QY 1292 GAAATCTCAGAAAGAGTTTACACACTTCCGAACATCTCGATGAAGAGTTCGCTCGCC 1351
Db 1368 GACAGTACGAGAACGAGGTCTACCGTCTGCTTAAGGTTCTCGACGAAAGGTGGACGCA 1427
QY 1352 TCACCTCGATCTCTGATGTCACCTTACAAAGCTTACAGAAAGAGGCTGACTACA 1411
Db 1428 TCACGCTTGAAGCTCTCGGCGGTGAGTCAAGGAGTCAAGGAGGAGGCTGAGTACA 1487
QY 1412 TCAAGCTTCAAGTTCAGGAGGCTTACAGTCTGATGCTTACCGTTATTA 1461
Db 1488 TCGGCGTTGAGCTTTCAGGCGCAATCAAGCGGAGCACTACCGCTACTAA 1537

RESULT 7
AX244105 1557 bp DNA linear PAT 28-SEP-2001
LOCUS
DEFINITION Sequence 97 from Patent WO0166573.
ACCESSION AX244105
VERSION AX244105.1 GI:15859169
KEYWORDS
SOURCE
ORGANISM Corynebacterium glutamicum
Corynebacterium glutamicum
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Corynebacteriaceae; Corynebacterium.
1
Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O., Haberhauer, G.,
Kim, J.W., Lee, H.S. and Hwang, B.J.
Corynebacterium glutamicum genes encoding metabolic pathway
proteins
JOURNAL Patent: WO 0166573-A 97 13-SEP-2001;
FEATURES BASF AKTIENGESSELLSCHAFT (DE)
source Location/Qualifiers
1..1557
/organism="Corynebacterium glutamicum"
/mol_type="genomic DNA"
/db_xref="taxon:1718"
101..1537
/feature="note=RXN00132"
/codon start=1
/transl table=1
/protein_id="CAC88531.1"
/db_xref="GI:15859170"
/translation="MAQVDMFKVDLSAEAGRHRQLRLAEYMPGLMQLRKEFADBP
LKGARIAGSIHMTVQAVLIETLTALGAEVWASCNIFSTQDEAAAAIIVGSGTVEP
```

```
AGVVFPAWKESLEBYWNCINQIFSWGDELPNNMILDGSDATMAVIRGREYEQALVP
PAPANSDDEVIAFLGMLREVLAAPBKGWKEAVKVTETTTGTVHRLVHFAEBGV
PFPAMVNDVATKSKFDNKNQTRSLDGNRATDMLMGKNVLVCGGVGVGKCAEA
PDGCGARVKTBADPINALQALMDGVSVTVDEAIEDADIVTATGKNKDIISPEOMLK
MKHALLGNIGHFDNEIDHSLHRDDVFTTTPQVDFEFTFTGSRSLIIVLSGRLLN
LGNATGHPFVMSNSFADQTTAQIQLPEHYR"
QY TELTKBQAEYIGVDVAGPFKPEHYR"
BASE COUNT 348 a 440 c 431 g 338 t
ORIGIN
Query Match 40.8%; Score 536.4; DB 6; Length 1557;
Best Local Similarity 64.3%; Pred. No. 6.1e-129;
Matches 923; Conservative 0; Mismatches 501; Indels 6; Gaps 2;
QY 32 TCGAGTACAGAAATGCGCATCAACCTCCATGTTCTCGGCGCTTAAGGAATTTACCCCTTG 91
Db 114 TGGACTTCAAGTTGCGGATCTTTCACTAGCAGAGGACGAGCTCACAGATTGCTTTG 173
QY 92 CTGAGAGAAATGCGCAGGTCTTATGTTCTTGTGAGGTTATTCGCTCTCTAAGCAT 151
Db 174 CAGAGTATGAGTCCAGGTCTCATGCAATGCGCAAGGAATTCGACAGAGAGCGCTT 233
QY 152 TGAAGGTTGTCAGAACTCTGTGTTCCCTCCACATGACAGTCCAGACAGCGCTCTCATCG 211
Db 234 TGAAGGCGCCGAAATGTTGTTCTTATCCATGAGCTCCAGACCGCGGTGCTTATTG 293
QY 212 AGACATCTCAGCTCTTGTGCTGATGTCAGATGGGCTTCTTGCACATCTTCTCTACAC 271
Db 294 AGACCTCTCACTGTTTGGCGCTGAGGTTCGTTGGGCTTCTTGCACATTTTCTCCACC 353
QY 272 AGATACAGCGCTGCTGTATCTGTTCTGTCGCCCAACAGGCACACAGAGAGCGCG 331
Db 354 AGATAGGCTGCGAGCGGCTATCGTTGTCGC---TCCGCGCACCGTCCGAGAGCGAGCTG 410
QY 332 GTATCCCACTCTTCGCTGCGCTGAAAGGCGGAAACACATCCACAGATACTGGGAGAACATACC 391
Db 411 GGTGTTCCAGTATTCGCTGGAAGGCTGATCTCACTGGAGGAGTACTGGTGTGCATCAACC 470
QY 392 GCGCTCTCACATGCGCAGATGTCAGGCCCAACAGCAGGTTGTCAGATGATGGTGGTATG 451
Db 471 AGATCTTCAGCTGGGCGGATGAGC---TGCCAAACATGATCTCTCGACAGCGCGGTGACG 527
QY 452 CTACACTCTCTCATCTCAAGGCTTCGAATTCGAACAGCGCGTCTCTCCAGAGCCAA 511
Db 528 CCACATGCGCTGTTATTCGCGGTGCGGAATACAGAGAGGCTGCTGTTCCACAGAGAG 587
QY 512 CAGAAAGCTGACAACTCGAATACCGCTCGCTTCTTGTCTACACTCAAGCAGGCTCTTCAACC 571
Db 588 AGGCCAACGATTCGATGAGTACATCGCATTTCTTGGCATGCTGCTGAGGTTCTTCTGTG 647
QY 572 AAGACAGAACCACTGGCACAGTTGCTGCGGCGCATGAAACGCTGTTCCGAAAGAGCAA 631
Db 648 CAGAGCTTGGCAAGTGGGCAAGATCGCTGAGGCGGCTTAAGGCTGTCTACCGAGAAACCA 707
QY 632 CAAAGGTTGCCACCGCTCTACAGCTTCGAGAGAGAGGAGGCAAACTCTCTTCCAGCCCA 691
Db 708 CCACCGGTGTCGACCGCTGTACGCTTCGCTGAGAGGCGTGTGCTTCCAGAGCA 767
QY 692 TCACGTCACAGAGCGCTGTTTAAAGTCCAAAGTTCGATTAACATCTACGCTGTCGCGCACT 751
Db 768 TGAAGCTCAACAGCGCTGTCAAAAGTCCAAAGTTGATTAACAGTACGCGCACCGCCACT 827
QY 752 CCGTTATCGATGATCAACAGGCTTCCGATGCTGATGATGCGGCGGAGAGACGCTCTCG 811
Db 828 CCGTATCGAGCGGATCAACCGCGCACTGACATGCTATGCGGCGGCAAGAACGCTGCTG 887
QY 812 TCATGGGTTACGGCGATGTCGCGCAAGGCTGCGCTCAATCCCTCCGTCGCGCAAGCGGCTC 871
Db 888 TCTGCGGTACGGGATGTCGCGCAAGGCTGCGCTGAGGCTTTCGACGCGCAGGCGGCTC 947
QY 872 GCGTTATCATCAGAACTCGACCAATCTGCGCTCTCCAGGCTGCCATGAAGGCTTACC 931
Db 948 GCGTCAAGGTCAACGAGGTGAGCCCAATCAACGCTCTTTCAGGCTCTGATGATGCTGCTACT 1007
```

QY 932 AGGTCCGGCCGATCGAGGAAGTGGTCAAGAGATGTCGATATCTTCGTTATCATCGACAGAA 991  
Db 1008 CTGTGGTCAACGGTTGATGAGCCATCGAGGACCGCATCGTCAATCACCGGACCGCA 1067  
QY 992 ACTGGGATATCATCTCTGTTGACATGATGGCCAGATGAAGAGTAAGGCTATGTCGGA 1051  
Db 1068 ACAAGGACATCATTTCTCTGGAGCAGATCTCAAGATCAAGAGATCAAGCTCTGTCGGCA 1127  
QY 1052 ACATCGGCACCTTCGATAACGAAATTGATACAGATGSCCTCATGAATATACCCAGGCATCA 1111  
Db 1128 ACATCGGCATCTTGAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1187  
QY 1112 AGACATCCCATCAAGCCAGAAATACGACATGAGGAAATCCAGATGCGCAGCTATCC 1171  
Db 1188 CCGCACACAGCATCAAGCCAGGATCGACGAGTTCACCTTCTCCACCGGTGCTCCATCA 1247  
QY 1172 TCTTCTTCTGAGGCGCGCTTCTTAACCTTGGCTGGCTAGAGTCAAGCTACCGCTTTTCG 1231  
Db 1248 TGTCTCTGTCGAGGTCGCTCTTGAACCTTGGCAAGCCACCGACACCCATCATTTG 1307  
QY 1232 TTATGTCAATGTCAATTCACAAACAGACACTCGCTCGAGCTCGACTCTACGAAAGAGAG 1291  
Db 1308 TCATGTCCAACTCTTTCGCGATCAGACCATTTGCGAGATCGAATCTTCCAAAGAGAG 1367  
QY 1292 GAATCTCGAGAGAGGTTTACACTTCGAGAGATCTCGATGAAGAGTGGCTGCGC 1351  
Db 1368 GACAGTACGAGAGAGGTTTACCGCTTGCCTTAAGGTTCTCGACGAAAGTGGCAGCA 1427  
QY 1352 TCCACCTCGATCTCTGATGTCACCTTACAAAGCTTACAGAGAGGCTGACTACA 1411  
Db 1428 TCCAGTTGAGGCTCTCGCGGTCAGCTCACCGAATGACCAAGGAGAGGCTGAGTACA 1487  
QY 1412 TCACGTTCCAGTTGAGGCTCTTACAGTCTGATGCTTACCGTTATTA 1461  
Db 1488 TCGGGTTGAGCTTGCAGGCGCCATTCAAGCCGAGCAGTACCGCTACTAA 1537

RESULT 8  
AX397890  
LOCUS AX397890 1939 bp DNA linear PAT 27-MAY-2002  
DEFINITION Sequence 1 from Patent WO0220806.  
ACCESSION AX397890  
VERSION AX397890.1 GI:21260756  
KEYWORDS Corynebacterium glutamicum  
SOURCE Corynebacterium glutamicum  
ORGANISM Corynebacterium glutamicum  
REFERENCE 1  
AUTHORS Farwick M., Huttmacher K., Brehme J., Pfeifferle W., Binder M.,  
Greissinger D. and Thierbach G.  
TITLE Nucleotide sequences which code for s-adenosyl homocysteinease  
(sahH)  
JOURNAL Patent: WO 0220806-A 1 14-MAR-2002;  
Degussa AG (DE)  
FEATURES  
source 1..1939  
/organism="Corynebacterium glutamicum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1718"  
227..1723  
/gene="sahH"  
227..1723  
/gene="sahH"  
/codon\_start=1  
/transl\_table=11  
/db\_xref="GI:21260757"  
/translation="MVRVKVFLIFILKGAQDMQVMDKVDLSIARAGRHLRL  
AYEYFGLMQLRKEPADEQPLKARIAGSIHMTQTAVLIEITLALGAEVWRASCHIF  
STQDEAAAIIVSGSVTERPAGVPVFAWKGSEUEFYWNCINQIFSWGDELPNMLDDG  
GDATWAVIRGREYEQAGLVPPAEANDSDVEYIAFLGMLRLVLAABEPKWKIABAVKGV

TEETTTGHRHLYHFAEBGVLPFFPMNVNDVATKSKFNDKYCTRHSLLIDGIRNATDMLM  
GGNVLVCCYGDVKGCAEAFDGGARVAVTEADPINALQALMDGYSVTVDEATEA  
DIVITATGNKDIISPEOMLAKWKDHALLGNIHFNEIDMSLHRRDDVTRTTIKPQVD  
EPTFTGSRILIVSEGRLNLGNATKHPSPFMSNSPADQTAQIELFQNEGYENEVY  
RLFKVLDEKVAIRIIVBALGGQLTFLTKQAEYIGVDVAGFPFKPEHRY"  
BASE COUNT 428 a 559 c 530 g 422 t  
ORIGIN  
Query Match 40.8%; Score 596.4; DB 6; Length 1939;  
Rest Local Similarity 64.5%; Pred. No. 61e-129;  
Matches 923; Conservative 0; Mismatches 501; Indels 6; Gaps 2;  
QY 32 TCGAGTACAGAAATGCGACATCAACCTCCATGTTCTGGCCGTGAAGAACTTACCTTG 91  
Db 300 TCGACTTCAAGGTTGCCGATCTTTCATAGCAGGACGAGCGTCAACAGATTCTGCTTG 359  
QY 92 CTGAGAGGAAATGCGAGCTCTTATGTTCTTCTGAGCGTTATTCGGCTTCTAAGCCAT 151  
Db 360 CAGAGTATGATGCGCAGCTCTCATGAGTTGCGCAGGAATTCGCAAGCAGCCTT 419  
QY 152 TGAAGGTTGTCAAAATCTCTGTTCCCTCCATGACAGTCCAGACAGCGCTCTCATCG 211  
Db 420 TGAAGGCGCCGCAATTGCTGTTCTATCCATGACGCTCCAGACCGCGTCTTATTG 479  
QY 212 AGACACTCAGCTCTTGTGCTGATGTGATGAGGCTTCTCTGCAACATCTTCTCTACAC 271  
Db 480 AGACCCCTCACTGCTTTGGGCGCTGAGGTTCTGTTGGGCTTCTCTGCAACATTTCTCCACC 539  
QY 272 AAGATACAGCGCTGCTGCTATCTGTTGCGGCCAACAGGCACACAGAGAGCCAGCGC 331  
Db 540 AGATGAGGCTGCGAGCGCTATCTGTTCTCGC---TCCGGCACCTCGAAGAGCAGCTG 596  
QY 332 GTATCCAGCTCTTGGCTGGAAGGCGAACAACCTCCAGAAATCTGGGAGAAACATATCC 391  
Db 597 GTGTTCCAGTATTCGCGTGAAGGCTGAGTCACTGAGGAGTACTGTGTGTCATCAAC 656  
QY 392 CGGCTCTCACATGCGCAGATGTTCAAGGCCACAGCAGGTTGTCGATGATGATGATGATG 451  
Db 657 AGATCTTCACTGGGCGGATGAGC---TGCCAAACATGATCTCTCGACGCGGTTGACG 713  
QY 452 CTACACTCTCTATCTCAAGGCTTTCGAATTCGAAACAGCGGCTGCTTCTCCAGAGCCAA 511  
Db 714 CCACATGGCTGTTATTCGCGTTCGGAATACGAGCAGGCTGCTGTTGTTCCACGAGAG 773  
QY 512 CAGAGCTGACAACTCGAATACCGCTGCTTCTTCTGCTACACTCAGCAGSTCTTCAAC 571  
Db 774 AGGCCAACGATTCGATGATACATCCATCTTGGGCATGCTGCTGAGGTTCTTCTGCTG 833  
QY 572 AAGACAAAGAACCTGCGACACAGTGTGCTCGCGCATGAACGGTGTTCGCAAGAGACAA 631  
Db 834 CAGAGCTTGCAGTGGGCGAAGTCTGAGGCGGTTAAGGTTCTCAACGAGAAACCA 893  
QY 632 CAACAGTGTTCACCGCTCTTACAGCTTCGAGAGGAGGCGAACTCTCTTCTCCAGGCCA 691  
Db 894 CCACCGTGTGCACCGCTCTTACCACTTCGCTGAAGAGGCGTGTGCTTCTCCAGCGCA 953  
QY 692 TCAAGCTCAACGAGCTGTTCAGGTCGAAAGTCCAAAGTTCGATACATCTACGGCTGCGCCACT 751  
Db 954 TCAAGCTCAACGAGCTGTTCAGGTCGAAAGTTCGATACATCTTCCAGGTCGCGCCACT 1013  
QY 752 CCCTTATCGATGATCAACCGCTGCTTCGATGTCGATGTCGCGGCAAGACAGCTCTCG 811  
Db 1014 CCCTGATCGAGGCGATCAACCGCGCTGATGATGCTCATGAGGCGCAAGAGCTCTTG 1073  
QY 812 TCATGGTTACGGCGATGTCGGCAAGGCTGCGCTCAATCCCTCGGTGGCCAGAGGCTCTC 871  
Db 1074 TCTCGGTTACGGCGATGTCGGCAAGGCTGCGCTGAGGCTTTCGAGGCGCAGGCGCTC 1133  
QY 872 CGCTTATCATCAAGAACTCGACCAATCTGCGCTCTCCAGGCTGCCATGAAGGCTTACC 931  
Db 1134 CGCTCAAGGTCACGAGCTGACCCATCAACCTCTTTCAGGCTCTGATGATGATGATG 1193  
QY 932 AGGTCCCGCATCGAGGAGTTCGTCAGGATGTCGATATCTTCTGTTACATGACACAGAA 991

D	1194	CTGTGTCACCGTTGATGAGCCATCGAGGACGCGCATCGTGATCACCGGACCGGCA	1253
Q	992	ACTGCGATATCATCTCTGTTGACATGATGGCCGACGAGGATAGGCTATTGTCGGTA	1051
D	1254	ACAAGGACATCATTTCTCTCGACGATGCTCAAGATGAAGGATACGCTCTGCTGGGCA	1313
Q	1052	ACATCGGCCACTTCGATAAGAAATGATACAGATGGCCTCATGAAATACCCAGGCAATCA	1111
D	1314	ACATCGGTCACCTTTGATAATGAGATCGATATCGATTCCTGTTGCACCGGACGAGGTCA	1373
Q	1112	AGCATATCCATATCAGCCAGATACAGATGTGGAAATCCAGATGGCCAGCTATCC	1171
D	1374	CCGACACACAGATCAAGCCACAGTGCAGGAGTTTCACTTCTCCACCGTCTGCTCCATCA	1433
Q	1172	TCCTCTTCTGTCGAGGGCGCTTCTTAACTTGGTGCGCTACAGGTCAACCATCTTTCG	1231
D	1434	TGCTCTGTCGAGGTCGCTGTTGAACCTTTGGACAGCCACCGGACACCATCATTTG	1493
Q	1232	TTATGTCATATGTCATCAAAACACGACACTCGCTCAGCTCGACCTCTACGAAAGAGAG	1291
D	1494	TCATGTCCAACTCTTTCCCGGATCAGACCAATTCGCGAGATCGAACTGTTCCAAAACGAAG	1553
Q	1292	GAAATCTCGAGAGAGGTTTACACACTTCCGAGCATCTCGATGAGAGTCCGCTGCC	1351
D	1554	GACAGTACGAGAACGAGGTCTACCGTCTGCTTAAGTTCTCGACGAAAGGTGGCAGCA	1613
Q	1352	TCCACCTCGGATCTCTCGATGCCACCTTACAACTTACACAGACGAGGCTGACTACA	1411
D	1614	TCCACCTGAGGCTCTCGCGGTTCAGTCACTCAGCACTGACCAAGGAGGCTGAGTACA	1673
Q	1412	TCACGTTCCAGTTCAGGTCCTTACAGTCTGATGCTTACCGTTATTAA	1461
D	1674	TCGGCGTTGAGCTTCAGGCGCCATTCAGCGGAGCACTACCGCTACTAA	1723
RESULT	9		
LOCUS	AP005276	332050 bp	DNA linear BCT 08-AUG-2002
DEFINITION	Corynebacterium glutamicum ATCC 13032	DNA, complete genome, section 3/10.	
ACCESSION	AP005276	BA000036	
VERSION	AP005276.1	GI:21323419	
KEYWORDS			
SOURCE	Corynebacterium glutamicum ATCC 13032		
ORGANISM	Corynebacterium glutamicum ATCC 13032		
REFERENCE	1		
AUTHORS	Nakagawa, S.		
TITLE	Complete genomic sequence of Corynebacterium glutamicum ATCC 13032		
JOURNAL	Unpublished		
REFERENCE	2	(bases 1 to 332050)	
AUTHORS	Nakagawa, S.		
TITLE	Direct Submission		
JOURNAL	Submitted (24-MAY-2002) Satoshi Nakagawa, Kyowa Hakko Kogyo Co. Ltd., Tokyo Research Laboratories; 3-6-6, Asahi-machi, Machida, Tokyo 194-8533, Japan (E-mail: snakagawa@axnagen.com, Tel:81-44-829-3031, Fax:81-44-813-1651)		
COMMENT	This sequence is conducted by collaboration of Kyowa Hakko Kogyo Co. Ltd. And Kitasato University.		
FEATURES	Location/Qualifiers		
source	1. .332050		
	/organism="Corynebacterium glutamicum ATCC 13032"		
	/mol_type="genomic DNA"		
	/strain="ATCC 13032"		
	/db_xref="taxon:198627"		
	/note="ATCC 13032"		
gene	314. .2518		
	/gene="Cg10654"		
CDS	314. .2518		
	/note="PF02554:Carbon starvation protein Ceta"		

gene	2522. .2725		
CDS	2522. .2725		
	/gene="Cg10655"		
	/codon_start=1		
	/transl_table=11		
	/product="Hypothetical protein"		
	/protein_id="BAB98048.1"		
	/db_xref="GI:21323421"		
	/translation="MMALTHALMKIPRAVWYLTLMGDTAYSKYVVLKHHHPDAP		
	PTREYWRKAYADODANPGARCC"		
	complement(3092..3703)		
gene	/gene="Cg10656"		
CDS	complement(3092..3703)		
	/gene="Cg10656"		
	/codon_start=1		
	/transl_table=11		
	/product="Hypothetical protein"		
	/protein_id="BAB98049.1"		
	/db_xref="GI:21323422"		
	/translation="MSNVINLVLMQASHQNVNLMNPSSDATDQILATFSSLLCCSQ		
	NRNIVSWGVPRSEERMEIKLTHRSASVSKSPFGTITQWENECTTIGNWDL		
	HRTGRANSTQHGSLHRCITVDVADHFLSVRSNLNMGVGNASDLVSVLLVRLFR		
	RTVGHFVNVHGNVTPSQRTNVAHTNSLKSNGVCHCKSCVIL"		
gene	3750. .4706		
CDS	/gene="Cg10657"		
	3750. .4706		
	/gene="Cg10657"		
	/codon_start=1		
	/transl_table=11		
	/product="Uncharacterized protein involved in propionate		
	catabolism"		
	/protein_id="BAB98050.1"		
	/db_xref="GI:21323423"		
	/translation="WMNTIYQALGOALHTTTATRSRKGAISSWKAPAFAPAKRMSIE		
	AVDRMGEPSPMEIGEDGVIAWLLSGLDHIYTIPLPAEGEAKRAILDYTKHSA		
	EYQSQAPIDLRSPMEKLAQGLDSDVSIVLHSHHTYVIGTSGNDPQKDFDAS		
	RETLDSINVIYFAVLEDRHRSYAPARHRETELWNKISTVEDEPWTYRHS		
	VDAKAFACARAVITFKDGTVEDELAADAPHPARPAPQVIOKFTLAEGVSE		
	KEODRFDAAQCTHELDSELNIELDILAKAPVIEGLF"		
	4715. .5629		
gene	/gene="Cg10658"		
CDS	4715. .5629		
	/gene="Cg10658"		
	/codon_start=1		
	/transl_table=11		
	/product="FAP phosphonotase and related enzymes"		
	/protein_id="BAB98051.1"		
	/db_xref="GI:21323424"		
	/translation="MFSAPATERRRKALRAALAAPEIARMPGAFSPLAARAIQEA		
	EGVYVSGAVVAADLALPDIGLITLTVFAHRSQIARVTDLPVLVDADTGTGPNAA		
	TVSELEDAGVACGHLEQVQKSGHLDGKGVGTDIMVRIAAVNERDEQFVICA		
	RTDAAGVEGSDAITERKAYADAGAMITFEALYSPADPFKFAANDIPELLANTFEG		
	KTELLPAQLLEDIGYNAVIYFVTLIRIAMGQVEQALGDIANTGTDWDRMQHRSRL		

YELLYNEYNAFDQOVFTYSADSYKPIF"  
5658. 6809  
/gene="Cg10659"  
5658. 6809  
/gene="Cg10659"  
/EC number="4.1.3.7"  
/note="PF00285:Citrate synthase"  
/codon\_start=1  
/transl\_table=11  
/product="Citrate synthase"  
/protein\_id="BAB98052.1"  
/db\_xref="GI:21323425"  
/translation="MSSATTIVRKGLGVYADYAVSKVMPETNSLYRQYAVEDIV  
ENCSEEVYLLHGLSPTAQQAELFNGRSYRSLDAGLILHSFKEAHFDMVR  
TAVSMGTQSEYFTDSEHRIKVGHTLLAQLPWLMIDIRRKGLDIIADPSGKSYVA  
ENLMSVGTGSPSPADVRDEKSLILYAEHFNASTFARVITSTKSDVYSAI  
TCAICALGPHGAEFVWHTMLAIDPNKAAWNNALDNKVNWVGFGHRYKGD  
SRVPSWKSRELAABHDCGKVMYENMRDAMDARTGLKPNLDGPAPVHLLGPV  
DFTFLFVIAKVGWTAHVEQYENNSLIRFSLSYNGESQREVAFIKR"  
complement (6846. .7040)  
/gene="Cg10660"  
complement (6846. .7040)  
/gene="Cg10660"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein"  
/protein\_id="BAB98053.1"  
/db\_xref="GI:21323426"  
/translation="MVPDVSQDVDELHPSFFELYADQAALDAHRCAHPFYWREAVA  
KHLVPGSQKNMVGTRMLHNG"  
complement (7129. .8172)  
/gene="Cg10661"  
complement (7129. .8172)  
/EC number="1.1.1.82"  
/note="PF02615:Malate/L-lactate dehydrogenase"  
/codon\_start=1  
/transl\_table=11  
/product="Malate/L-lactate dehydrogenases"  
/protein\_id="BAB98054.1"  
/db\_xref="GI:21323427"  
/translation="MPEVTVNAQQLTVLCTDILTKTVPAADAHLVGDSLVQADLMGH  
PSGVLRLPMVTRHSGAMTHAHVEKLVLDGLAVLADGHNGIGOVLDHARKEAVT  
RAMMGIVASVRNSHEFTAMVYTRKAAOGCVSILTTNAPAPWAGREKRGITPN  
PNSIAPGETATVQDANTAVARKLYHAEQTNMIPETMAITSEGHTPDRAHIN  
GVVLPMAGHKYALSENMVLSVLTGSQHSYKVGHPYDTPPGAGHLFIALLDVAAP  
RDPQDFDALSDLVGVKVPKPAQNTTEIFYPGSESDRAHRKNSAHGISLPKTMWEL  
QELAIENHVVTHR"  
complement (8265. .8987)  
/gene="Cg10662"  
complement (8265. .8987)  
/gene="Cg10662"  
/note="PF00392:Bacterial regulatory proteins, gntR family"  
/codon\_start=1  
/transl\_table=11  
/product="Transcriptional regulators"  
/protein\_id="BAB98055.1"  
/db\_xref="GI:21323428"  
/translation="MVKDAQSPATRKADQVYQTPRIEDGTLNPGQMSVYWLVE  
HTGAKTRVDRALRADELILEPQAPWVSPSLRHILKOLFERRIVEVALEEI  
SVGAKSPRIEGETSLAADFRELENSADDAFTADFRILTSKFDPLVAANTHNOFLG  
RSILSKPHTRLRIIAHSDHARLRQSVQHEIEMCEAVASGDLRSAGAACRQHLLHVE  
KSLITLAINADSTGSGQIDIRS"  
complement (9094. .9798)  
/gene="Cg10663"  
complement (9094. .9798)  
/gene="Cg10663"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical membrane protein"  
/protein\_id="BAB98056.1"  
/db\_xref="GI:21323429"  
/translation="MRFSSRVLPALLITAVPIPTASAAITLADTKELCIASINTDSSA

VVTFMNSIEDSVRQRLDEDAQDPGIIKAAIESYIAQDNPATAEALQVRLDAISGE  
GLAMLLPDDPTLADPNAESESFKTEYTYDEAKDIISGFSSDPASDVLSQQQAATTGTR  
TAIRAEVFPADRTDDYNESTALKEDFQNCIDAIDARPIQLOVILIGNAIALAVIL  
GIRAWTNSRKQSKHSQ"  
complement (9878. .12094)  
/gene="Cg10664"  
complement (9878. .12094)  
/gene="Cg10664"  
/EC number="1.1.1.42"  
/note="NIGR00178:monomer\_idh: isocitrate dehydrogenase,  
NADP-dependent"  
/codon\_start=1  
/transl\_table=11  
/product="Monomeric isocitrate dehydrogenase (NADP+)"  
/protein\_id="BAB98057.1"  
/db\_xref="GI:21323430"  
/translation="WAKIWIWTRTDEAPLLATYSLKPVVEAFAATAGIEVETRDISLAG  
RIIAQFPELATEQKGNALAEIGELAKTPEANIKLENISASVPOLKAAIKELQDQ  
YDIPELNDATIDBEKOLLARYNAVKGSPVNLREGNSDRRAPIAVKNFVKFFHRM  
Query Match 40.8%; Score 596.4; DB 1; Length 332050;  
Best Local Similarity 64.5%; Pred. No. 7e-129;  
Matches 923; Conservative 0; Mismatches 501; Indels 6; Gaps 2;  
Qy 32 TCGAGTACAGAATTGCCGACATCAACCTCCATGTTCTCGGCCGTAAAGAACTTACCTTGG 91  
Db 120723 TGGACTTCAAGGTGGCGATCTTCTACTAGCAGGACGAGCGTACACAGATTGCTTGG 120782  
Qy 92 CTGAGAAGGAAATGCCAGAGCTCTTATGTTCTTCGTGAGCGTTATTCGGCTTCTTAAGCCAT 151  
Db 120783 CAGAGTATGAGATGCCAGGTCTCATGCAGTTGCGAAGGAATTCGACAGCAGCAGCTT 120842  
Qy 152 TGAGGGGTGTCAGAACTCTGTTCCCTCCACATGACAGTCCAGACAGCGTCTCTCATCG 211  
Db 120843 TGAAGGCGCCGGAATGCTGGTTCTATCCACATGACGCTCCAGACCGCGCTGCTTATGG 120902  
Qy 212 AGACACTCACAGCTCTTGGTGTGATGTACAGATGGGCTTCTCGCAACATCTTCTACAC 271  
Db 120903 AGACCTTCACTGTTTGGCGCTGAGGTTGTTGGGCTTCTCGCAACATTTCTCCACCC 120962  
Qy 272 AAGTACAGCGGTGCTGTCTATCGTTGTCGCCCAACAGGCACACAGAGAAGCCAGCGG 331  
Db 120963 AGGATGAGGCTGCAGCGGCTATCGTTGTCGCG---TCCGECACCGCTCGAAGAGCAGCTG 121019  
Qy 332 GTATCCAGCTTTCGCTCGAGGCGGCAACACTCCCAAGATACCTGGGAGAGACATACC 391  
Db 121020 GGTATCCAGTATTCGGGTGGAAGGCTGAGTCACTGAGAGGATCTGGTGGTGCATCAAC 121079  
Qy 392 GCGCTCTCACATGGCCAGATGTCAGGCGCCACAGCAGGTTGTCGATGATGGTGGTATG 451  
Db 121080 AGATCTTCAGCTGGGCGGATGAGC---TGCCAAACATGATCTCGACGACGCGGTGACG 121136  
Qy 452 CTACACTCTCATCTCCAGGGCTTCGAATTCGAAACACCGCGTCTGTTCAGAGCCAA 511  
Db 121137 CCACCATGGCTGTTATTCGCGGTGCGGAATACAGCAGGCTGGTCTGGTTCACACGAG 121196  
Qy 512 CAGAAGCTGCAACACTCGAATACCGCTTCTGTATCACTCAAGCAGAGCTTCTCAAC 571  
Db 121197 AGGCCACAGATTCCGATGAGTACATCGCATTCCTGGCGATGCTGGTGAGGTTCTGCTG 121256  
Qy 572 AAGACAGAAACCACTGGCACACAGTGTCTGCGCGCATGAACGGTGTTCGCGAGAGACAA 631  
Db 121257 CAGAGCCTGGCAAGTGGGCAAGATCGCTGAGGCCGTTAAGGGTGTCAACGAGAAACCA 121316  
Qy 632 CACAGGTGTCCACCGCTCTACAGCTCCGAGAGAGGCGCAACTCTCTCCAGGCCA 691  
Db 121317 CCACCGGTGTGCACCGCTCTGTACACTTTCGTGAAGAGCGGTGCTCTTCCAGCGA 121376  
Qy 692 TCACGCTCAACGACGCTGTTTACAAAGTCCAAGTTCGATTAACATCTACGGCTGCGCGCACT 751  
Db 121377 TGAACGCTCAACGACGCTGTCCACAGTCCAAGTTTGATTAACAAGTACGCGCACCGCCACT 121436  
Qy 752 CCCTTATCGATGGTATCAACCGTGTCTCCGATGTCTATGATCGGCGGCAAGACAGCTCTCG 811

Db 121437 CCCTGATCAGCGCATCAACCGGCCACTGACATGCTCATGGGCGCAAGACGTCCTTG 121496  
Qy 812 TCATGGGTTACGGGATGTGGCAAGGGTGGCTCAATCCCTCGTGGCGCAAGCGCTC 871  
Db 121497 TCTGGGTTACGGGATGTGGCAAGGGTGGCTGAGCTTCGACGCGCGGCGCTC 121556  
Qy 872 GCGTTATCATCAGAACTCGACCAATCTGGCTCTCCAGCTGCGCATGGAAGGCTACC 931  
Db 121557 GCGTCAAGGTCAACGAGCTGACCAATCAACGCTCTTCAGGCTCTGATGGATGGCTACT 121616  
Qy 932 AGGTCCGCGCATCGAGGAGTGTCAAGGATGTCGATATCTTCGTTTACATGACAGAA 991  
Db 121617 CTGTGTACCTTGATGAGGCAATCGAGAGCGCGCATCTGATACCGGACCGCA 121676  
Qy 992 ACTGCGATCATCTCTGTTGACATGATGCGCCAGATGAAGATPAAGGCTATTGTGGTA 1051  
Db 121677 ACAAGGACATCAATTCCTTCGAGCAGATGCTCAAGATGAAGATCAAGCTCTGCTGGCA 121736  
Qy 1052 ACATCGGCACTTCGATACGAAATGATACAGATGGCTCATGAATACCCAGGATCA 1111  
Db 121737 ACATCGGTCACATTCGATTAAGATGATGATATCCCTGTTGCAACCGCGACGACGTCA 121796  
Qy 1112 AGCAGATCCCATCAAGCCAGATACGACATGTGGGAATCCAGATGGCCACGCTATCC 1171  
Db 121797 CCGGACACAGATCAAGCCAGGTCGAGGTTCACTTCTCCACGGTCTCCATCA 121856  
Qy 1172 TCTTCTTCGTGAGGCGGCTCTTTAACTTGGCTGGCTACAGGTCAACCATCTTTTCG 1231  
Db 121857 TCGTCTGTTCGAGGTCGCTGTGTAACCTTTGGCAACGCCACCGGACACCCATCAATTG 121916  
Qy 1232 TTATGTCATGTCATTCAGAAACGACACTCGCTCAGCTCGACTCTAGAAAGAG 1291  
Db 121917 TCAATGTCACATCTTTCGCGATGACCAATTCGCGATCGAATGTTCCAAACGAG 121976  
Qy 1292 GAAATCTCGAAGAAGGTTTACACACTTCCGAACATCTCGATGAAGAGTTCGCTCGCC 1351  
Db 121977 GACAGTACGAGAACGAGTCTACCGTCTGCTTAAGTCTTCGACGAAAGGTGGACGCA 122036  
Qy 1352 TCCACCTCGATCTCTGATGTCACCTTACAACTTACAGAGGACGAGGTCGACTACA 1411  
Db 122037 TCCACGTTGAGGCTCTCGGCGGTGAGTCAACGAACTGACCAAGGACGAGGTCGATACA 122096  
Qy 1412 TCAACGTTCCAGGTCGAGGTCCTTACAGTCTGATGCTTACCGTTATTA 1461  
Db 122097 TCGGCTTGAGCTTCAGGCCCATTCAGCCGAGCACTACCGCTACTTA 122146

RESULT 10  
AX127145  
LOCUS AX127145 349980 bp DNA linear PAT 11-MAY-2001  
DEFINITION Sequence 7061 from Patent EP1108790.  
ACCESSION AX127145 AX114121  
VERSION AX127145.1 GI:14041133  
KEYWORDS  
SOURCE Corynebacterium glutamicum  
ORGANISM Corynebacterium glutamicum  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
1  
REFERENCE Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,  
Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.  
TITLE Novel polynucleotides  
JOURNAL Patent: EP 1108790-A 7061 20-JUN-2001;  
KYOWA HAKKO KOGYO CO., LTD. (JP)  
FEATURES Location/Qualifiers  
i. .349980  
/organism="Corynebacterium glutamicum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1718"  
/note="Seq 1 to long (3,309,400) split in 11, seq 7061  
0.600.001 0.949.980",  
BASE COUNT 79274 a 90638 c 88277 g 81341 t

ORIGIN  
Query Match 40.8%; Score 596.4; DB 6; Length 349980;  
Best Local Similarity 64.5%; Pred No. 7e-129;  
Matches 923; Conservative 0; Mismatches 501; Indels 6; Gaps 2;  
Qy 32 TCGAGTACAGAAATTCGCGACATCAACCTCCATGTTCTTCGCGCGTAAAGAACTTACCCCTG 91  
Db 188673 TGGACTTCAAGTTTGGCGATCTTTCATAGCAGAGGAGGAGGAGTCCACAGATTGCTCTTG 188732  
Qy 92 CTGACAGGAAATTCGCGAGTCTTATGTTCTTCTGAGCGTTATTCGCTTCTTAAGCCAT 151  
Db 188733 CAGATATGAGTTCGCGAGTCTCATGCGATTGCGCAAGGAATTCGACAGCAGAGCCCTT 188792  
Qy 152 TGAAGGTTGTGAGAAATCTCTGTTTCCCTCCACATGACAGTCCAGACAGCCGTCCTCATCG 211  
Db 188793 TGAAGGCGGCCGAAATGCTGTTCTATCCACATGACGCTCCAGACCGCGTGTATTG 188852  
Qy 212 AGACACTCAGAGCTCTTGTGCTGATGTCAGATGGGCTTCTGCAACATCTTCTTACAC 271  
Db 188853 AGACCTTCACTGTTTGGGCGCTGAGTTCTGTTGGGCTTCTGCAACATTTCTTCCACC 188912  
Qy 272 AAGATACAGCGCTGCTGCTATCTGTTGCGGCCAACAGSCACACAGAGAGCCAGCG 331  
Db 188913 AGGATGAGCTGACGCGCTATGTTGTCGCGC---TCCGSCACCGTCCGAGGCAAGCTG 188969  
Qy 332 GTATCCAGCTTTCGCTCGAAGGCGAAACACTCCAGAACTCTGGGAGAACATATCC 391  
Db 188970 GTGTTCCAGTATTCGCTGGAAGGCTGAGTCACTGGAGGACTGTTGTTGTCATCAACC 189029  
Qy 392 GCGCTCTACATGCGCAGATGTCGAGGCCACAGAGGTTTCTGATGATGTTGTTGATG 451  
Db 189030 AGATCTTCAGCTGGGCGATGAGC---TGCACAAACATGATTCCTCGACACGCGGCTGACG 189086  
Qy 452 CTACACTCTCTATCTCAAGGGCTTCGAAATTCGAAACAGCGGCTGCTGTTCCAGAGCAA 511  
Db 189087 CCACCATGGCTGTTATTCGCGCTCGCGAATAGAGCAGGCTGTTGTTCCACGACAG 189146  
Qy 512 CAGAACTGACACACTCGAATACCGCTCGTTTCTGCTACATCTCAAGCAGGTTTCAACC 571  
Db 189147 AGGCGCAAGATTCGATGATGATCGCATTCGCGATTCGCGTGAAGTCTTCTGCTG 189206  
Qy 572 AAGACAGAACCACTGGCACACAGTTCGCGGCAAGACGCTGTTCCGAGAGCAA 631  
Db 189207 CAGAGCTGGCAGAGTGGGCAAGATCGCTGAGGCGCTTAAGGTTGTCACGAGGAACCA 189266  
Qy 632 CAACAGTGTTCACCGCTCTACAGCTTCGAGAGGAGGCGCAAACTTCCTTCCAGCCCA 691  
Db 189267 CCACCGGTTGACCGCTGTACCACTTCGCTGAAGAGGCGTGTGCTTCCACGCA 189326  
Qy 692 TCACGCTCAGAGCGCTGTATCAAAAGTCCAGTTCGATACATCTACGGCTGCGGCCACT 751  
Db 189327 TGAAGCTCAACGACGCTGTACCAAGTCCAAAGTTTATTAACAGTACGCGCACCGCCACT 189386  
Qy 752 CCCTTATCGATGTTATCAACCGCTTCGCGATGATGATTCGCGCGCAAGACACTCTCG 811  
Db 189387 CCCTGATCGAGCGCATCAACCGCGCACTGACATGCTCATGCGCGGCAAGACGCTGTTG 189446  
Qy 812 TCATGGGTTACGCGATGTCGCAAGGCTGCGCTCAATCCCTTCGCTGGCCAAAGCGCTC 871  
Db 189447 TCTGCGGTTACGCGATGTCGCAAGGCTGCGCTGAGGCTTTCGACGCGCAGGCGCTC 189506  
Qy 872 GCGTTATCATCAGAACTCGAGCCCAATCTCGGCTCTCCAGGCTGCCATGGAAGGCTACC 931  
Db 189507 GCGTCAAGGTCACGAAAGTGCACCAATCAACGCTCTTCAGGCTCTGATGATGGCTACT 189566  
Qy 932 AGGTCCGCGCATCGAGGAAGTTCGTAAGGATGTCGATATCTTCGTTACATGCAAGAA 991  
Db 189567 CTGTGTCACGTTGATGAGGCCATCGAGAGCGCGACATCTGATCACCGGACCGGCA 189626  
Qy 992 ACTGCGATATCATCTCTGTTGATCATGATGCGCCAGATGGAAGGCTATTGCTGTTA 1051  
Db 189627 ACAAGGACATCTTCTTCGAGCAGATGCTCAAGATGAAGATCACGCTCTGCTGGCA 189686

Db	182	AGACCTCTACTGCTTTGGGGCGCTGAGGTTCTGTTGGGCTTCTTGCAACAATTTTCTCCACCC	241
Qy	272	AAGATACAGCCGCTGCTGCTATGCTTATGTTGTCGGGCCAACAGGCACACAGAGAAGCCAGCGG	331
Db	242	AGGATGAGGCTGACGGGCTATCGTTGTCCGG---TCGGGCACGTCGAAGAGCCAGCGT	298
Qy	332	GTATCCCACTTCTGCTCGCTGGAAGGGCAACACTCCAGANACTCGGAGAACACATACC	391
Db	299	GTGTTCCAGTATTCCGTGGAAGGGTAGTCACTGGAGGAGTACTGTGTGTGCACTAACCC	358
Qy	392	GCGCTCTCACTGCGCAGATGGTCAAGGCCCCACAGCAGGTTGTGCATGATGTTGTGTGATG	451
Db	359	AGATCTTCAGCTGGGGCGATGAGCT--GCCAAAATGATCATCTCGACGACGGGTGAGC	415
Qy	452	CTACACTCTCTCATCTCCAAAGGGCTTCGAATTCGAACAACGCGGTGCTGTTCAGAGCCAA	511
Db	416	CCACCATGCGCTGTTATTCCGCGGTGCGGAATACGACAGGCGTGTGTGTTCACACAGCAG	475
Qy	512	CAGAAGCTGACAACTCGGAATACGCTGGGTTCTTGCTACACTCAAGCAGGTTCTTCAACC	571
Db	476	AGGCCAACGATTCGATGAGTACATCGCATTTCTTGGGCATGCTCGCTGAGGTTCTTGTGTC	535
Qy	572	AAGAACAAGAACCTATGCGACACAGTTGCTGCGCGGCAATGAACGGTGTTCGAAAGACAA	631
Db	536	CAGAGCTGGCAAGTGGGGCAAGATCGCTGAGCGCGTTTAAAGGTGTCAACGGAGAAACA	595
Qy	632	CAACAGGTGTCACCGCGCTCTACCAAGTTCGAAAGGAGGGCAAACTCTCTCTCCAGACCA	691
Db	596	CCACGGTGTGACCGCGCTGTACCACTTTCGCTGAAGAAGGGCTGCTGCTTTCCACAGCA	655
Qy	692	TCACGTCAACGACGCTTTTAAAGTTCGAAGTTCGATAACATCTACGGCTGCGGCCACT	751
Db	656	TGAAGGTCACAGACGCTGTCAACAGTTCGAAGTTTGATAAACAAGTACGCGCACCGCCACT	715
Qy	752	CCCTTATCGATCGTATCAACCGTCTTCGATGTCTATGATCGCGGCAAGACAGACTCTCG	811
Db	716	CCCTGATCGAAGGCACTCAACCGCGCACTGAATGCTCATGGCGGCAAGAGCTGCTTG	775
Qy	812	TCATGGGTTACGCGCATGTCCGCAAGGGGTGCGCTCAATCCGCTCCGTGGCAAGCGGCTC	871
Db	776	TTCGCGTTACGGCGATGTCCGCAAGGGCTGCGCTCAGGCTTTTCAGCGCAGGCGGCTC	835
Qy	872	GCGTTATCATCACAGAACTCGACCCCAATTCGCGCTCTCCAGGTCGCAATGGAAGGCTACC	931
Db	836	GCGTCAAGGTTCACCAAGCTACCCCAATCAACGCTCTTCAGGCTCTGATGGATGGCTACT	895
Qy	932	AGGTCGCGCGCATCGAGGAAGTCTGTCGAAGATGTCGATATCTTGTTACATGACACAGAA	991
Db	896	CTGTGTCACCGTTGATGAGGCCATCGAGGACGCCGACATGCTGATCACCGCACCGGCA	955
Qy	992	ACTCGGATTCATCTCTGTTGACATGATGCCCGAGATGAAGGATTAAGGCTATTGTGCGTGA	1051
Db	956	ACAAGGACATCAATTTCCCTTCGACAGATGCTCAAGATGAAGGATCAAGCTCTCTGGGCA	1015
Qy	1052	ACATCGGCGACTTCGATAACGAATTCGATACAGATGGGCTCATGAATACCCAGGCATCA	1111
Db	1016	ACATCGGTCACTTGTATATGAGATCGATGCAATCCCTGTTGACCGGACGACGTCA	1075
Qy	1112	AGCACAATCCCAATCAAGCCAGAAATACGACATGTGGGAATTCACAGATGGCCACGCTATCC	1171
Db	1076	CCGCGACACGATCAAGCCACAGGTGACGAGTTACCTTTCTCCACCGTCTGCTCATCA	1135
Qy	1172	TCCTTCTTGTGAGGGCGCTTCTTAACTTGGCTGCGCTACAGGTCAACCATCTTTGCG	1231
Db	1136	TCGTCCTGTCCGAAGTCTGCTGTTGAACCTTGGCAAGCCACCGCACCCCATCATTTTG	1195
Qy	1232	TTATGTCAATGTCTATTTCACAAACAGACACTCGCTCAGCTCGACTCTACGAAAAGAGAG	1291
Db	1196	TCATGTCCAATCTTTTCCGCGATCAGACCATTTGGCGAGATCGAATGTTTCCAAAACGAG	1255
Qy	1292	GAATCTCGAAGAAGGTTTACACATTCGGAAGCATCTCGATGAAGAAGTGCCTCGCC	1351
Db	1256	GACAGTACGAGAACGAGTCTACCGTCTGCTGCTAAGGTTCTCGAGAAAAGTGGCAGCA	1315







Db 1256 GACAGTACGAGAACGAGTCTACCGTCTGCTTAAGGTTCTTCGACGAAAGGTGGCAGCA 1315  
QY 1352 TCCACCTCGGATCTCTGATGTCACCTTACAAAGCTTACACGAGCGGCTGACTACA 1411  
Db 1316 TCCACCTGAGGCTCTCGGCGGTGAGTCTACCGAAGTACGACGAGGAGGCTGAGTACA 1375  
QY 1412 TCAACGTTCCAGTTCGAGGCTCTTCAAGTCTGATGCTTACCGTTA 1457  
Db 1376 TCGGCTTGAGCTTCGAGGCGCCATTCAGCGGAGCAGTACCGCTA 1421

RESULT 13  
AP005216 308750 bp DNA linear BCT 20-DEC-2002  
LOCUS Corynebacterium efficiens YS-314 DNA, complete genome, section 3/11.  
DEFINITION AP005216 BA000035  
VERSION AP005216.1 GI:23492422  
KEYWORDS  
SOURCE Corynebacterium efficiens YS-314  
ORGANISM Corynebacterium efficiens YS-314  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
1  
REFERENCE  
AUTHORS Kwarabayasi, Y., Yamazaki, J., Hino, Y., Kikuchi, H., Nakamura, Y., Ikeo, K., Suzuki, M., Mashima, J., Itoh, T., Yamagishi, A., Nishio, Y., Ueda, Y., and Sugimoto, S.  
TITLE The entire genomic sequence of Corynebacterium efficiens YS-314  
JOURNAL Published Only in Database (2002)  
REFERENCE 2 (bases 1 to 308750)  
AUTHORS Kwarabayasi, Y., Yamazaki, J., Hino, Y., Kikuchi, H. and Director-General of Biotechnology Center.  
TITLE Direct Submission  
JOURNAL Submitted (17-MAY-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; Nishihara 2-49-10, Shibuya-ku, Tokyo 151-0066, Japan (E-mail: biocenter.go.jp, Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)  
COMMENT Kwarabayasi, Y. is officially affiliated with the National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, 305-8566 Japan  
Nakamura, Y., Ikeo, K., Suzuki, M. and Mashima, J. are at the National Institute of Genetics, Mishima, Shizuoka, 411-8540 Japan  
Itoh, T. is at the Japan Biological Information Research Center, Koto-ku, Tokyo, 135-0064 Japan  
Yamagishi, A. is at Tokyo University of Pharmacy and Life Science, Hachioji, Tokyo, 192-0392 Japan  
Nishio, Y., Ueda, Y. and Sugimoto, S. are at the Ajinomoto Co., Inc., Kawasaki, Kanagawa, 210-8681 Japan  
The other authors are at the National Institute of Technology and Evaluation, Shibuya-ku, Tokyo, 151-0066 Japan.

FEATURES  
source  
1..308750  
/organism="Corynebacterium efficiens YS-314"  
/mol\_type="genomic DNA"  
/strain="YS-314"  
/db\_xref="taxon:196164"  
155..598  
/note="CE0586, similar to AB007159-8|AAK47889.1| percent identity: 63 in 145 aa"  
/codon\_start=1  
/transl\_table=11  
/product="putative 50S ribosomal protein L13"  
/protein\_id="BAC17396.1"  
/db\_xref="GI:23492423"  
/translation="MSYHFPSKSDITRWYVIDATVLCRLATHAALLRGKGLP  
ANVGDGHVLIINADVAITSNKDRNTRHSGYEGGLKNTGRSLRLEHPERVLE  
ESTIGMPHNKLITAAVSKLVHFAAGSEHPYAAQPEYIEKKVQAQ"  
598..1146  
/note="CE0587, similar to AL583918-48|CAC29873.1| percent identity: 59 in 159 aa"  
/codon\_start=1  
/transl\_table=11  
/product="putative 30S ribosomal protein S9"

CDS  
155..598  
/note="CE0586, similar to AB007159-8|AAK47889.1| percent identity: 63 in 145 aa"  
/codon\_start=1  
/transl\_table=11  
/product="putative 50S ribosomal protein L13"  
/protein\_id="BAC17396.1"  
/db\_xref="GI:23492423"  
/translation="MSYHFPSKSDITRWYVIDATVLCRLATHAALLRGKGLP  
ANVGDGHVLIINADVAITSNKDRNTRHSGYEGGLKNTGRSLRLEHPERVLE  
ESTIGMPHNKLITAAVSKLVHFAAGSEHPYAAQPEYIEKKVQAQ"  
598..1146  
/note="CE0587, similar to AL583918-48|CAC29873.1| percent identity: 59 in 159 aa"  
/codon\_start=1  
/transl\_table=11  
/product="putative 30S ribosomal protein S9"

CDS  
1282..2670  
/note="CE0588, similar to AL031317-14|CAA20392.1| percent identity: 62 in 448 aa"  
/codon\_start=1  
/transl\_table=11  
/product="putative phospho-sugar mutase"  
/protein\_id="BAC17397.1"  
/db\_xref="GI:23492424"  
/translation="WSEPIQNEVSNVADADIAATAATEFTNTIGASTSTEE  
ETEPALVDGPWTVGRKRAIVRVVAGSGEITCNGETLEDFFPNKLQOLIK  
FLVILREGQFDIFANLGGGGTQAGAFRLAALNAYNPAERPTLKKAGFLIRDA  
RAVERKKAGLHKARRAPOYSKR"  
1282..2670  
/note="CE0588, similar to AL031317-14|CAA20392.1| percent identity: 62 in 448 aa"  
/codon\_start=1  
/transl\_table=11  
/product="putative phospho-sugar mutase"  
/protein\_id="BAC17398.1"  
/db\_xref="GI:23492425"  
/translation="MWCRAAGRWTCITMLGNTLFTDGVRLANKTUTAPLALKLGA  
AAHLVAGTRPHRRPVAIVGRDPRVSGEMLAALAAAGSRGVDLRVGVPTPGVA  
PLTDYDADMGWVLSASHNMPDNGIKFESAGGHLKDEVEDERVMODLPEEGPTG  
HGIGRVIEAPDARGVLOHLADAVPTDLSGITVVVDAANGAASVIAPOAYEAAKAV  
IAHNTFNAYINERKCSITHMOQIQAALVHAGDGLAHAGDADRCCLAVDSGNLVDG  
DQIMALAIAMKENSELKNLTATVMSNLGLKAMKADIQRTTKVGRVIRLEELN  
AGGFALGEGSGHIVLPDHGTGTGDTLITGLSLMARMAATGKPLSELAAAMTVLPQVLI  
NVFVADKSSIMKSANVQAAVAEEELGCTGRVILRPSGTTELFRVWVAABQBOARR  
VACRLAAVVAAE"  
2742..3071  
/note="CE0589"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAC17399.1"  
/db\_xref="GI:23492426"  
/translation="WLRGIVNSGLNVRGALTARTMLCEATHORORHEVAPPAPV  
SAAGQGFSTWGAQVATWLELHGVGDKRLDAVEATADAAREVGVFRDADAGIGRLG  
QVDEGR"  
3068..4462  
/note="CE0590"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAC17400.1"  
/db\_xref="GI:23492427"  
/translation="MSASISYGSAYVSQLQAASRTYPPGSIPIERVLKAYTITGLD  
TDQLRGSRSAAGSGETGGLLGLKPLGRMGVAGGMIINELVDVRSWVNRNEAE  
EVAAREADAADITITRESDDGVVAILGQAGIIAETISNHLSTDPVSHPEAFIATV  
QACAEIINDAAAVILGCGDRDRAIESCYGALIEHGREVCBAPAGVAVAGVGSGGG  
AGTIPSPVPTVTTSATGATAPADVPTADSPACTATGCTEACDCPPGQDEK  
FVSGPADEDETCGSPFPEDDWEHTITGASIGIGLIVGVGALVDFLIGVAAQ  
TWGDAAMPVPSFPAAPLAATVPEVPAPAPAPAPAPAPAPAPAPAPAPAPAPAP  
EPPKPPQGMVQAAATAVQEPYVQVYQAAATPAAPVAPPTAAAPAGNSGDA  
AGDAPGRVRRVGGW"  
4456..4716  
/note="CE0591"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAC17401.1"  
/db\_xref="GI:23492428"  
/translation="WVAVEVFAVQFRORTAARMAEPEKALAKAQCEVAKTSPGEPG  
RROAASPPQVTPASVARTPQQPGRGRGRGPVQSILRRA"  
complement(4743..5663)  
/note="CE0592, similar to ALS83918-51|CAC29878.1| percent identity: 44 in 254 aa"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="BAC17402.1"  
/db\_xref="GI:23492429"  
/translation="NAHFGRIILPSGVTLSTILDVAENLKKHLSKLSKRGPHRVWVG  
DLVYAGLPKIVTPAEGNGVGVAFGHDWMDPKIKHYHOTLHFLASWGIAPAPNTETG  
FLPHRGFAADLSDALQILAGVKLGAGTKVGNPGLGVGSGGVAVLAANAKAV  
RAYGLVPASTSPSDAAAPVAPAGLIITGTADLGMFDAGDPKAVKAWGGDVYREL  
DNGNQGFSEDTPKLVGVGRPQTAGQELARGLLTGILLQLAGERSYKASDFEAE  
AKKVTYWGIDLEKAPQDESPIPLNSN"  
5737..6840  
gene

CDS	<pre>/gene="alr" 5737..6840 /gene="alr" /EC number="5.1.1.1" /note="CE0593, similar to AX063847-1 CAC25165.1  percent identity: 68 in 362 aa" /codon_start=1 /transl_table=11 /product="alanine racemase" /protein_id="BAC17403.1" /db_xref="GI:23492430" /translation="MNLITRIDLDIAHNTRLLKNRVGAALKMAVVKADYVHGMDR VAPVWAANGADFGVATIAEALALRATGITPILCWISFQDMWAAVEAGVIDLAVIS PHARVLVADPTTSRAIRSVKVDTHLRSGVDQWDVFTLLDCHDGLVFLGFLTH LSCADEFNETDQETFRRAIDARALGLEVPENHLCNSPATLTPDLHMDVRFPG VALYGLRPILGLDHGRPAMTWAGAITVVKPIKRGTSYGLTMRABADGFVAVPAG YADGPRAAQDHLRVHUNGHDYPOVGRVMDQFVIFLGDNPHGVTAGDEAVIFGGTGM SATELADALATINYEVCPTGRTVRYDTGEGK" 6837..7340 /note="CE0594, similar to AL583918-56 CAC29985.1  percent identity: 50 in 156 aa" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="BAC17404.1" /db_xref="GI:23492431" /translation="MKPDPFASGTRRLDTHAQTAQFGBELGRALEAGVVVLDPGLGA GKTFPTQLARGNQVGRVTSPTFVIAEHRHSVGGPTLIHLDAYRLIGBDPDDVSD PTGALSDLDLTDLTAVVVAEWGGVLVEQITTSYLLISIDRETAVTDPESBARIFT WOMRSH" 7340..9132 /note="CE0595, similar to AX066613-1 CAC26534.1  percent identity: 41 in 530 aa" /codon_start=1 /transl_table=11 /product="putative transport protein" /protein_id="BAC17405.1" /db_xref="GI:23492432" /translation="MELFLAANPLIAVLAVLGLAIGRISLFGVSLGAANLIVLV VSTLNPDIQIPAFVQGLAMFVYVIGISGAPAFFRFSRGWKLITWILLVLSUTA LAWLILAFGLAGAGAFAGSLSTPGMAAVVMDPAQGVFVIGYSIAYPGVLL LSLIVAAVGLAKLVNHTDAREGLVTEPLVWKVRIIGESITIGLPLSGQIIL ATRVIPHEHRLADLTPIKPELMELVINGTVNADRAIALAGGECDFKDEHLVYS RTVSNPDIQRTVAELDPVANGEMARIQGDTEIVPHRDVLYNVSORVVAAPR NGEVRPLGSEKALGDVNLPLPAIGLSIGLLGALPVLPGDTMWLGGGSPVAG LLGLANRTGPIWQVLPFRANRITISITLGLALFVAGVTGAGAGFQALTDQSFVNG VGLAITVSLVCAVGVGMMLKLKWDSEMGVAAGATINPAIISYLNQDTGPDLANRGY ATVYPTAMIGKILACQVLFLL" 9183..10529 /note="CE0596, similar to AX065751-1 CAC26115.1  percent identity: 79 in 430 aa" /codon_start=1 /transl_table=11 /product="putative prolyl aminopeptidase" /protein_id="BAC17406.1" /db_xref="GI:23492433" /translation="MARPGGTWPDWGLTGGGFVSTRGNLDGMTELSTHLELR DLBTLTVDADRTIDIPARIATPFGGTELPVLTFLQGGNENAPRPSLEPLNP SWLKAEHVHVMVLDQRTGLSTPVSRLLEBHTTAEIYVLSHLRADGIVRDCEN RYLGVARNVVLGSGGGTTLHLSTHPDLDLDFVITGLSADRPAEDVYANCYR MYNSEQYRPFEDRVGIQRLVERAGETIIPTEGVVSESRLSLGHLGLDNL DIHLLERPRSNFLDLADLPFGARNPLYVYVHSSYADGVVTDWAAQRYVPEAF REDPTLLGFVPEWADTVFALRPWKDVALAAQEWPKLYDAAALSTSGATGAAY YANDVFVPLDPSLETAHPLFGVKLYITSGEHNGLRASGRVLAHLIDLAHGRKVR" 10536..11210 /note="CE0597, similar to AX065799-1 CAC26139.1  percent identity: 71 in 224 aa"</pre>	Db	196018	GACTTCAAGGTGCGGATCTGTCCCTCGTGAAGCGGTGCGCACCATCGCTCGCG	196077
		Qy	94	GAGAAAGAAATGCCAGGTCCTATGGTTCCTGAGCGGTATTCGGCTTCTTAAGCAATG	153
		Db	196078	GAATACAGATGCGGGGTCTGATGACGTGCGAGGAGTATGCAGAGGAGGACCAATG	196137
		Qy	154	AAGGGTGTGAGAATCTCTGGTTCCCTCCACATGACAGTCCAGAGAGCCGTCCTCATCGAG	213
		Db	196138	AAGGTGCGCGCATCGGGGTCCATTCATGACGGTGCAGACCGCGGTCTCTGATCGAG	196197
		Qy	214	ACACTGACAGCTCTGGGTGCTGATGTCAGATGGCTTCTGCAACATCTTCTACAGAA	273
		Db	196198	ACCTGACCGCTCTCGGTGCGGAGGTCCGCTGGGTCTTCGCAACATCTTCTCCACCGAG	196257
		Qy	274	GATACAGCGCTGCTGCTATCGTTGTGCGGCCCAACAGGACACACAGAGAGCCAGCCGCT	333
		Db	196258	GACGAGGCGCTGCGGCCCATCGTCTCGG---TGACGGGACCCCGGAGGACCCCGAGGT	196314
		Qy	334	ATCCCACTTTCGCTTGAAGGGGCGAAACATCCCAAGATATCTGGGAGAAACATACCGC	393
		Db	196315	GTTCGGGTGTTCGCTTGAAGGGGTGAACCCCTGGATGATGATCTGCTGTCATCAACAG	196374
		Qy	394	GCTCTCATATGCGCAGATGCTCAAGGCCACACAGAGTGTGCGATGATGCTGATGCT	453
		Db	196375	ATCTTCAGCTGG---GAGGGTGAATCTCCGAACATGATCTCTGATGACGAGGTGACGC	196431
		Qy	454	ACACTCTCATCTCCAAAGGGCTCGAATTCGAAACAGCGGTGCTGTTCAGAGCAACA	513
		Db	196432	ACCATGCGCTTATCCGTGGACGTGAGTAGAAGCGGGCGTCTGACCGCAGCGGAG	196491
		Qy	514	GAACTGACAACTCGAATACCGTGGTCTTCTGCTACACTCAAGCAGAGTCTTCAACCA	573
		Db	196492	GCCATGACTCTCGATGATGATACATCGCTTCTCGGATGCTCCGCGAGGTCTTGCGCGAG	196551
		Qy	574	GACAAAGAACCACTGGCACACAGTGTCTGCGCGCATGAAAGGTGTTTCGAGAGAGAA	633
		Db	196552	GAGCGGACAAAGTGAGACCGCGCTCGCGACTCCATCAAGGGGTGTCCAGGAGGACACC	196611
		Qy	634	ACAGGTCTCCCGCTCTACAGCTCGAGAGGAGGGAACACTCTTCTCCAGCGCTC	693
		Db	196612	ACCGGTCTCCCGCTCTGTATCACTTCCGCGAGGAGGGGTGCTGCCCTTCCCGCGCATG	196671
		Qy	694	AAGCTCAACAGCGCTTCTACAAAGTCAAGTTCGATFAACATCAAGGTGCGCGCACTCC	753
		Db	196672	AATGTCAATGATCGGTGACCAAGTCCAAATTCGACACACAGTACGGTACCGCGCACTCC	196731
		Qy	754	CTTATCGATGGTATCAACCGTGTCTTCGATGTATGATCGCGCGCAAGACAGTCTCGTC	813
		Db	196732	CTGATCGATGGCATCAACCGTCCACCGGACATGCTCATGGGTGTGAAGAACGTCTGGTC	196791
		Qy	814	ATGGGTTCAGCGGATGTGCGCAAGCGGTGCGCTCAATCCCTCGTGGCGAAGCGCTCGC	873
		Db	196792	TGTGGCTACCGTGTGCGCAAGCGGTGCGCGAGCTTTCGACGCGCAGGTGTCACT	196851
		Qy	874	GTATATCATCAGAACTCGACCCCAATCTCGGCTCTCCAGGTGCGCATGGAAGGTACCGAG	933
		Db	196852	GTCCGGGTCAACGAGGCGGATCCCATCAATGCTGACGCACTGATGGATGATATCTCC	196911
		Qy	934	GTCCCGCGCATCGAGGAAGTGTGCAAGATGTGATATCTTCTGTTACATGACAGAAAC	993
		Db	196912	GTGGTCAACCGTTCGACGAGGACCATCGCGCGCGGACATCGTCATCAGCGGCACCGGCAAC	196971
		Qy	994	TGCGATATCATCTCTCTTTCGATGATGGCCGACATCAAGGATTAAGGTATTGTTCGGTAAAC	1053
		Db	196972	RAGGACATGATCTCTTATGACGAGATGCTCAAGTGAAGGATCAGCGCTGCTGGGCAAC	197031
		Qy	1054	ATCGGCCACTTCGATAAGAAATTGATACAGATGGCTCATGAAATACCGAGGATCAAG	1113
		Db	197032	ATTGGTCACTTCGACAAATGAGATGACATGCAATTCCTCTGTCACCGCGATGATGTCATC	197091
		Qy	1114	CACATCCCAATCAAGCCAGAAATGACATGTGGAAATTCGCCAGATGGCCAGGTATCTTC	1173

Query Match 38.8%; Score 567.2; DB 1; Length 308750;  
Best Local Similarity 63.3%; Pred. No. 4.8e-122;  
Matches 904; Conservative 0; Mismatches 518; Indels 6; Gaps 2;  
Qy 34 GAGTACAGATTCGCGACATCAACCTCCATGTTCTCGGCCCGTAGGAACTTACCTTCTCT 93

197092	CGACCAACATTAAGCCGCGAGTGAGTGTCACCTTCCCCAAACGCGAAGTCCCATCATC	197151
QY	CTTCTTGCTGAGGCGCGCTCTTAACTTGGTGCCTACAGTCAACCATCTTTTCGTT	1233
197152	GTGCTGTCCGAGGAGCTCTGTCTAACTGGGCAACGCCACCGCCACCCAGCTCTCGTC	197211
QY	ATGTCAATGTCATTTCACAAACAGACACTCGCTCAGCTCGACCTCTACGAAAGAGAGA	1293
197212	ATGTCCACTCTCTTGGCCGACGACGATGGCCAGATCGAACTGTTTCCGAAAGAGGCG	197271
QY	AATCTCGAAGAAGGTTTACACACTTCGAAGCATCTCGATGAAGAAGTCTCGTCGCTC	1353
197272	CAGTACGAAACCAAGTCTACCGTCTGCCAAGATCCTCGATGAGAAGTTCGCCGATC	197331
QY	CACCTCGGATCTCTCGATGTCACCTTACAAAGCTTACACAGAGCAGGCTCACTCATC	1413
197332	CAGTGGAGGCCCTCGTGGAGAGCTACCGAGCTGACGAGACAGGCCGAGTACATC	197391
QY	AACGTTCCAGTTGAGGTCCTTACAAGTCTGATGCTTACCGTTATTAA	1461
197392	GGTGTGGATGTCGCGGCTCGGTTCAAGCCGAGCACTACCGTACTGA	197439

RESULT 14	
AB088224/c	
LOCUS	AB088224 210614 bp DNA linear BCT 11-JUN-2003
DEFINITION	Streptomyces rochei plasmid pSLA2-L DNA, complete sequence.
ACCESSION	AB088224
VERSION	AB088224.1 GI:30698345
KEYWORDS	.
SOURCE	Streptomyces rochei
ORGANISM	Streptomyces rochei
REFERENCE	Bacteria; Actinobacteriya; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces. 1 Kinashi, H., Fujii, S., Hatani, A., Kurokawa, T. and Shinkawa, H. Physical mapping of the linear plasmid pSLA2-L and localization of the eryAI and actI homologs Biosci. Biotech. Biochem. 62, 1892-1897 (1998)
AUTHORS	
TITLE	
JOURNAL	

2	Suwa, M., Sugino, H., Sasaoka, A., Mori, E., Fujii, S., Shinkawa, H., Nimi, O. and Kinashi, H.
TITLE	Identification of two polypeptide synthase gene clusters on the linear plasmid pSLA2-L in <i>Streptomyces rochei</i>
JOURNAL	Gene 246 (1-2), 123-131 (2000)
MEDLINE	20231737
PMED	10767533
3	Hiratsu, K., Mochizuki, S. and Kinashi, H.
TITLE	Cloning and analysis of the replication origin and the telomeres of the large linear plasmid pSLA2-L in <i>Streptomyces rochei</i>
JOURNAL	Mol. Gen. Genet. 263 (6), 1015-1021 (2000)
MEDLINE	20408175
PMED	10954087

REFERENCE	4	Mochizuki, S., Hiratsu, K., Suwa, M., Ishii, T., Sugino, F., Yamada, K. and Kinashi, H.
AUTHORS		The large linear plasmid pSA2-L of <i>Streptomyces rochei</i> has an unusually condensed gene organization for secondary metabolism
JOURNAL		Mol. Microbiol. 48 (6), 1501-1510 (2003)
MEDLINE		22676966
PUBMED		12791134
REFERENCE	5	(bases 1 to 210614)
AUTHORS		Mochizuki, S., Hiratsu, K. and Kinashi, H.
TITLE		Direct Submission
JOURNAL		Submitted (15-JUN-2002) Haryasu Kinashi, Hiroshima University, Department of Molecular Biotechnology, Graduate School of Advanced Sciences of Matter; 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8530, Japan [E-mail: kinashi@hiroshima-u.ac.jp, Tel: 81-824-24-7869, Fax: 81-824-24-7869]
COMMENT		The nucleotide sequence has been determined by using restriction fragments and nested deletion fragments of the ordered cosmid library of pSA2-L.

pSLA2-L is 210,614 bp long with G+C content of 72.8 % and carries 143 ORFs.

Gene prediction was based on the unique codon usage in Streptomyces (aibb et al., Gene 30:157-66 (1984)) using the FramePlot program of Ishikawa and Hotta (FEMS Microbiol Lett 174: 251-253 (1999)) as implemented at <http://www.nih.go.jp/~jun/cgi-bin/frameplot.pl>. Where possible we chose an initiation codon (atg, gtg, ctg or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13 bp before the initiation codon). If this could not be identified we chose the most upstream initiation codon.

Three-fourths of the pSLA2-L DNA are occupied by biosynthetic genes (three PKS gene clusters for lankacidin, lankamycin and an unknown type II polyketide, a carotenoid biosynthetic gene cluster, many regulatory genes and others). The range of each biosynthetic gene cluster has been deduced by comparison with similar gene clusters in most cases and may be revised in future.

Location/Qualifiers

1. .210614  
 /organism="Streptomyces rochei"  
 /mol\_type="genomic DNA"  
 /strain="7434AN4"  
 /db\_xref="taxon:1928"  
 /plasmid="pSLA2-L"  
 /note="linear plasmid"

1. .1992  
 /note="left terminal inverted repeat, TIR-L; shows 99.4 % (1981/1992) sequence identity to TIR-R (complement (208623..210614))"

683. .2188  
 /note="N-terminal sequence is almost identical (435/437) with that of ORF143 at the right end of pSLA2-L until the inner end of TIR-L"

ORF1 (501 aa)  
 similar to A1590463 Streptomyces coelicolor putative helicase, SCP1.136 (879 aa); homology is seen until the inner end of TIR-L"

/codon\_start=1  
 /transl\_table=11  
 /product="putative helicase"  
 /protein\_id="BAC76459.1"  
 /db\_xref="GI:30698346"  
 /translation="MSTSRSTQRTAAQAEADVAVTALELPVRLAPERGLRTQVIMATYSGKTVAAKSAEKVGRVIVPSDLTLTQTEAKREAGRTGPMIGVSURGEDVAFPNITDEVDLWMPDKVTPVATYASLGTLTERAHGSLPCWDLIVDRAHRTSRLGKVAVVDHDKTRIPSLRVLYMTATPRLWQLDEDAEGCELVASWEDDPDLGLGACFTLTSEALDRGICAPQVVYVDITDTQLQAQLGVGCRSDVARGAALQTAALIKKASSENFRFTLVPHVMKAEAFAGLPDVAKRLHAAGCLGVPRTIWMVLCGGHKKRRRVLTGFTSGIATGTVVEKGLSVKVLGEGVDPRCDSDVYWDVYRGGKPLQVQVGRARIGCGGKVASVVDLDFGETADNMLTSPYNGLLAFRQAPLITGHEGFAARAPFGFVFRVPLVGRGAQRGASGAPSHSIPTPSANSLSTSWRAP"

/note="ORF2 (426 aa)  
 similar to AE004736-10 Pseudomonas aeruginosa hypothetical protein (442 aa)"

/codon\_start=1  
 /transl\_table=11  
 /product="putative secreted protein"  
 /protein\_id="BAC76460.1"  
 /db\_xref="GI:30698347"  
 /translation="MRRKPRALLIGPLGLVAVLAPAAHASADTYVGTWAAPTTAAADTSVYEDDQRTGTVHLSVAGDSLRFVFNFTGTSPLTIGVHAARPAAQGPATPAADSDRVRVFGTRSPATLFGTQRSDVPALPTTAGDGLVLSLYLPQRTPGSTVHSAAYQHNFAAGDVTGAPDLTVFVSTINTSWHFLSGSVDRRAGTADSSVVLGDSITGEGHTLSDRRWPDLLAGLRARDDVAGLVNAGIGNRLLRDPPDPGSAESAFAFYFGEALDKRIYATITPAGTITGYFTPREARVQVNDMTSGAPDTVLQDVALDRPAPLPHLLPAYDGGDGLHPNDGAMAAARAFPLDSL"

complement (3659..4645)  
 /note="ORF3 (328 aa)  
 similar to Y00459-2 Streptomyces griseus regulatory protein, StrR (350 aa)"

/codon\_start=1

```
/transl_table=11
/product="StrR-like regulatory protein"
/protein_id="BAC76461.1"
/db_xref="GI:30698348"
/translation="MTALAGNLDQETRSIDPSLVPGYSPRLCGEDAEHVARLADTD
EALPILVERSTLRVLDQHWRLVAKAGHTTIEVRFQGAEEAPFLAVSNWTHGL
PLSRDRAAQAQILADQPHLSDRVAGIAGTAKTVAALPLAAGTTPNPQARRRD
PIRDLDTGRKRAELLAQRPQASVREVARHAGISPATASVRRRLASGRSPVER
NTPGARPGGTGRGATPGGGAAPFVSVVRTPRPDSPLVRLRDPRLSRKSGRRL
LRLCCGAVERTALLAQTVPVPHCTDLVLAELAREYADLWAFAREVATDG"
complement(4922..5812)
/note="ORF4 (296 aa), lankacidin biosynthesis protein
similar to AE004623-4 Pseudomonas aeruginosa
pyrroloquinoline quinone biosynthesis protein B (304 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein B"
/protein_id="BAC76462.1"
/db_xref="GI:30698349"
/translation="WILLGTAAAGGFPQWNCALCARGRRGELPARSQEAVSGDG
RWLLNLSPIRTQLLAAPALTDPGPRDTPVRGILLTDAEVDHALGLVIRGATGL
TVYAAPVRGALSARELPVRLGDLRDYAPWDRDATAPGPAVAGGLTVTAHVPGTKAPK
YAHAPDPAWVCAYRIEDPATGGALVAPCLATWPDGPDLLASATCALLEDGPFPSA
GELGTATSSAGAGOSLMGLPVAGPGSSALAHARGLRIYTHLNNINPLDSSAA
HAAVREAGVEVLDGSELVL"
complement(5815..6909)
/note="ORF5 (364 aa), lankacidin biosynthesis protein
similar to A0277117-5 Gluconobacter oxydans putative pqqE
protein (359 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein E"
/protein_id="BAC76463.1"
/db_xref="GI:30698350"
/translation="MADPAVAGPAGMLIETHRCPLCPYCSNPLVRRBAELTCBQ
WDITQVREHVLQVQHFSGGFLARPLDPLVGHARPLRGAVYVNLVTSVGLTAEARH
DIARGVRLGSLQADPAAGATAGARVHTAKLEAARAVTAAGLPLTVNLVLRHN
IDRTGRMVDLADIGDRLELANTQVYCGNLENAALMPTAQLAAREAVRHARTEY
AGSPELVVADYIDRPRKPCNDGWSQLTVTPAGDVLPFCAAYAITTLVPEVALER
PUSLWYASRYNAYRTGWNREKTPCERHADHGGRCQAFQLTGDAANTDPACSL
SPHRLVDLAALAEVTDGPFVAFVPRGPVA"
complement(6902..7174)
/note="ORF6 (90 aa), lankacidin biosynthesis protein
similar to AL603642-200 Sinorhizobium meliloti putative
pyrroloquinoline quinone biosynthesis protein D (98 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein D"
/protein_id="BAC76464.1"
/db_xref="GI:30698351"
/translation="WTGLPEPTVPLRFGLVRLTRDPARGELALLPERVVVNLDTAAAV
LAHCGTTSIAGIVERLAEYEGVSAEDVRELLRLAQRVVDLHG"
complement(7171..7899)
/note="ORF7 (242 aa), lankacidin biosynthesis protein
similar to AE004625-5 Pseudomonas aeruginosa
pyrroloquinoline quinone biosynthesis protein C (250 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein C"
/protein_id="BAC76465.1"
/db_xref="GI:30698352"
/translation="WMSVTVREVAAPWSEAFRQLHLESSYWDHPHFRMRHEGLL
DEGELTWAANRWYQCLPQKADTAVANCPLEVRPQWLSRIVVHDGADACAGAK
WLRLAEAVLRDSDVHDERLVLAGTFRAVDVAFARRPWLEMAASGLTLPSPGLL
AHLRLGSLREYEWIAERGFYFTARIEVVGPEGRSLDLVARHANSRQEQACVRLA
PKCRVLNVLASLDYHTNGATRS"
complement(7926..8030)
/note="ORF8 (34 aa), lankacidin biosynthesis protein
similar to AL603642-197 Sinorhizobium meliloti putative
pyrroloquinoline quinone biosynthesis protein A (31 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein A"
```

```
/protein_id="BAC76466.1"
/db_xref="GI:30698353"
/translation="WRTSGKELPAKXWHRPDPFTIDTGMVTAIYFSR"
8420..10057
/note="ORF9 (545 aa), possible lankacidin resistance
protein"
CDS
Query Match 38.6%; Score 563.8; DB 1; Length 210614;
Best Local Similarity 63.8%; Pred. No. 3e-121;
Matches 916; Conservative 0; Mismatches 492; Indels 27; Gaps 3;
Qy 33 CGAGTACAGAAATTCGCGACATCAACCTCCATGTTCTCGCGCGTAGAGAACTTACCTTCG 92
Db 101625 CGACTTCAAGGTCCGACCTGTCCCTGGCGGAGTTCGCGCGAAGAGATCACTTCG 101566
Qy 93 TGAGAAGAAATGCCAGGTCTTATGTGTTCTTCGTGAGCGTTATTCGCGCTTAAAGCAT 152
Db 101565 CGAGCACAGATCCCGGCTGATGGCGATCCGCAAGAGTAGCGCGGGGAGCGCGCT 101506
Qy 153 GAAGGGTGTGAGAAATCTTGTTTCCTCCATGATGAGTCCAGACAGCCGCTTCCTATCGA 212
Db 101505 GCGCGGCGCGGATCACCGGCTCCCTGCACATGACCGTGCAGACCGCGCTGCTGATCGA 101446
Qy 213 GACACTCACAGCTCTTGTTGCTGATGTCAGATGGGTTCTCTGCAACATCTTCTCTACACA 272
Db 101445 GACCTGTGTCCTCGCGCGGAGTCCGCTGGGCTCTTGCACATCTTCTCCACCA 101386
Qy 273 AGATACAGCGCTGTGCTGTTGTCGCGCAACAGGACACACAGAGAGAGCCAGCGG 332
Db 101385 GSAACACGCGCGCGCCATCGCGCTCGCGGACGCGCGGACACACCTCGGG 101326
Qy 333 TATCCAGTCTTCCTGGAAGGCGAAACACTCCAGATCTGGGAGACACATACCG 392
Db 101325 GCTCCCGGTCTTCCTGGAAGGCGAGCCCTGGAGGAGTACTGGTGTGCACCGAGCA 101266
Qy 393 CGCTCTCACATGCCAGATGGTCTC-----AAGGCCACACAGCAGGTGTTCGATGATGTTGG 446
Db 101265 GCGCTGACCTGCGCGGACAGCCACCGCGGCGCGGACATGATCTCTGGAGAGCGCG 101206
Qy 447 TGATGTACATCTCTCATCTCCAGGGTTCGAAATTCGMAACAGCGGTGCTGTTCAGA 506
Db 101205 CGAGCCACCTCTCTGCTCCACAGGGGCTGAAATACCGCAAGACCGGT-----GA 101155
Qy 507 GCGAACAGAGCTGACAACTCGATACCGTGGCTTCTGCTACACTCAAGCAGGTCTT 566
Db 101154 GTGCGCGCGCGCGACACAGAACTGGCGCGCGCTGGCGCGCCCTGCTGAAAGCG----- 101100
Qy 567 CAACCAAGACAAGAACCACTGGCACACAGTGTGTCGCGGCATGAACGGTGTTCGGAAGA 626
Db 101099 -----AGCAACCTCGACTGGACCGGCATCGCTCCAGATCCCGGGGTGACGGAGGA 101047
Qy 627 GACACACAGGTGTCCACCGCTCTACCGCTCGAGAGAGGAGGCGAACTCTCTTCC 686
Db 101046 GACACAGCGGGTGTGACCGGTGTGACAGATGACAGCTGACGGTGTGCTGTGTTCCC 100987
Qy 687 ACCCATCAAGCTCAAGACGCTGTTTACAAAGTTCGAATTCGAATCACTTACGGCTGCG 746
Db 100986 GSCGATCAATGTGAACGACGCGTGAAGAGTTCGAATTCGACAAAGTAGTAGGGTGC 100927
Qy 747 CCACTCCCTTATCGATGATCAACCGTGTCTCGATGTGATGATCGCGGCGGACAGACG 806
Db 100926 GCATTCGTGTGAGCGGATCAACCGTGTGCGGACGCGTGTGATCGCGGCGGAGACCG 100867
Qy 807 TCTGCTCATGGTTACGGCGATGTCGCGAGGGTGTGCTCAATCCCTCCGTGTGCCAAGG 866
Db 100866 TGTGTTGTGCGGTAGCGTGAITGTGCGCAAGGGTGTGCGGAGTCCCTGCGGGGCGAGG 100807
Qy 867 CGCTCGCGTTATCATCAGAACTCGACCCAACTCGCTCTCCAGGCTGCCATGGAAGG 926
Db 100806 CGCGGGGTGATCGTCAACGAGATCGACCCGATCTGCGCGCTGCGAGGCGCGATGAGCGG 100747
Qy 927 CTACAGGTTCGCGCGCATCGAGGAAGTGTGATATCTTCGTTCATCGTACATCGAC 986
```

Db 100746 TTATCAGTGAACGACGCTGGAAGAGGTCTCGAGACGGCGGACATCTTCGTACACAGAC 100687  
QY 987 AGGAAACTCGCATATCATCTCTGTGACATGATGATGCCAGATGAAGGATAAGCTATTGT 1046  
Db 100686 GGCACACAGGACATCATCTGCTCGACATGCGCAGATGAGCACAGGCGATCGT 100627  
QY 1047 CGGTAACTCGGCGACCTTCGATACGAATTTGATACAGATGGCTCATGAATACCCAGG 1106  
Db 100626 GGGGAACATCGCTCACITTCGACACAGAGATCGACATGCGCGGTCTGCGCGGTTCCGG 100567  
QY 1107 CATCAAGCACATCCCAATCAAGCCAGAAATACGACATGCGGAATTCACAGATGCCAGC 1166  
Db 100566 CATCGTCAAGGACGAGGTCAAGCCGACGCTCCACGTCGACGTTCCCGACGCGAAGT 100507  
QY 1167 TATCTCTCTTCTGTGTCGAGGCGCGCTTCTTAACCTTGGCTGCGCTACAGTCAACCATC 1226  
Db 100506 GATCATCGTGTCTCGAGGCGCGTCTGTGAACCTGGCAACCGCACCGGCCCGCTC 100447  
QY 1227 TTTCGTTATGCAATGTCATCAAAACAGACATCGCTGACGTGACCTCTACGAAA 1286  
Db 100446 GTTCGTTGATGCACTCTTCGCGACGACAGCGTGGCGAGATCGACTCTTTCACCA 100387  
QY 1287 GAGAGGAATCTCGAGAGAAGGTTTACACATCTCCGAAGCATCTCGATCAAGAGTGC 1346  
Db 100386 GCCGACAGTACCCGACCGCGGTGTACGCTGCCAGCACTGACGAGAAGTGC 100327  
QY 1347 TGCCTTCACCTCGGATCTTCGATGTCACCTTACAAAGTTACACAGAGGCTGA 1406  
Db 100326 CGTCTCCACTCGACGCGCTCGCGCTCAAGCTGACCAAGCTGCGCGCGGAGCAGCGC 100267  
QY 1407 CTACATCAACGTTCCAGTTCAGGTCCTTACAGGCTGTGATCTTACCGTTTAA 1461  
Db 100266 GTACATCGAGTTCAGGTCGAGGCGCGCTACAGCGGACCACTACCGTACTGA 100212

RESULT 15

AE012843/c

LOCUS AE012843 12968 bp DNA linear BCT 01-JUL-2002  
DEFINITION Chlorobium tepidum TLS section 64 of 194 of the complete genome.

ACCESSION AE012843 AE006470

VERSION AE012843.1 GI:21646659

KEYWORDS

SOURCE

ORGANISM

Chlorobium tepidum TLS

Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;

Chlorobium.

1. (bases 1 to 12968)  
Eisen, J.A., Nelson, K.E., Paulsen, I.T., Heidelberg, J.F., Wu, M.,  
Dodson, R.J., Deboy, R., Gwinn, M.L., Nelson, W.C., Haft, D.H.,  
Hickey, E.K., Peterson, J.D., Durkin, A.S., Kolonay, J.L., Yang, F.,  
Holt, I., Unavay, L.A., Mason, T., Brenner, M., Shea, T.P., Parksey, D.,  
Niernan, W.C., Feldblyum, T.V., Hansen, C.L., Craven, M.B., Radune, D.,  
Vamathevan, J., Khouri, H., White, O., Gruber, T.M., Ketchum, K.A.,  
Venter, J.C., Tettelin, H., Bryant, D.A. and Fraser, C.M.  
The complete genome sequence of Chlorobium tepidum TLS, a  
photosynthetic, anaerobic, green-sulfur bacterium  
Proc. Natl. Acad. Sci. U.S.A. 99 (14), 9509-9514 (2002)  
12091901

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

2. (bases 1 to 12968)  
Eisen, J.A., Nelson, K.E., Paulsen, I.T., Heidelberg, J.F., Wu, M.,  
Dodson, R.J., Deboy, R., Gwinn, M.L., Nelson, W.C., Haft, D.H.,  
Hickey, E.K., Peterson, J.D., Durkin, A.S., Kolonay, J.L., Yang, F.,  
Holt, I., Unavay, L.A., Mason, T., Brenner, M., Shea, T.P., Parksey, D.,  
Niernan, W.C., Feldblyum, T.V., Hansen, C.L., Craven, M.B., Radune, D.,  
Vamathevan, J., Khouri, H., White, O., Gruber, T.M., Ketchum, K.A.,  
Venter, J.C., Tettelin, H., Bryant, D.A. and Fraser, C.M.  
Direct Submission  
Submitted (30-APR-2002) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
unpublished

TITLE

JOURNAL

REMARK

FEATURES

source

Location/Qualifiers

1..12968

/organism="Chlorobium tepidum TLS"

/mol\_type="genomic DNA"  
/strain="TfLS"  
/db\_xref="taxon:194439"  
105..179  
/gene="trnA-Gln-2"  
105..179  
/gene="trnA-Gln-2"  
/product="trnA-Gln"  
305..1195  
/gene="mtd"  
/note="synonym: CT0719"  
305..1195  
/gene="mtd"  
/notes="identified by match to PFAM protein family HMM  
PF00763"  
/codon\_start=1  
/transl\_table=11  
/product="methylentetrahydrofolate dehydrogenase"  
/protein\_id="AAW1956.1"  
/db\_xref="GI:21646660"  
/translation="MKVLEPNPVAASFRDAVRROISEQLTINIVGILASDDPASITY  
ADYTRAGCDVGHFDLRKCEPSVRATLEAANRDSAVHGFVYPIWGKRDALRD  
LISPHDVGELSHWIKLVANERFDDTERREKSLPCTPLAITKLEVTEAVRGL  
PFGQDITLPNSESVCGRPLAYMLSNDRGVYSPDINGGVYDVSSDHESEPVTRBE  
ALQSQDIVITGVPSHPFKVRAELKPGALCLNFSYIQNFEPEAKEASLIIPRVGPM  
TVAMWRNALQIYNYHHEV"  
1185..1808  
/gene="CT0720"  
1185..1808  
/gene="CT0720"  
/notes="similar to PID:694120 SP:Q49135; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="serine cycle enzyme, putative"  
/protein\_id="AAW1957.1"  
/db\_xref="GI:21646661"  
/translation="MKSDFLGSASTYLDLSASADPTFGGAAAATVAAQGAALLS  
MVNLTIGKFPADVEHEVSIIDQCETARRQMLNDRNEVFTIMQIYKPKATP  
EEADARAAIQQALKASAEVFFLFRCLAMPLADRLKLEIKNPVLSVDVIGRVILY  
AGLSAMANVEVNLASIDDAEFCKAKREVMLPAIASMGVSWRGLAGV"  
complement (1910..3325)  
/gene="sahH"  
/notes="synonym: CT0721"  
complement (1910..3325)  
/gene="sahH"  
/notes="identified by match to PFAM protein family HMM  
PF00670"  
/codon\_start=1  
/transl\_table=11  
/product="adenosylhomocysteinase"  
/protein\_id="AAW1958.1"  
/db\_xref="GI:21646662"  
/translation="MTTEAAVLSDYKVADISLAENGRKEIEIAEKEMPGMATRKKEG  
KFLAGATAGSLHMTQITAVLIETLELGVADRWASCNIFSTQDHAALAAAGVPV  
FAMKGTLDVYVWCTQILPEFEGGLNLIYDDGDATLMHFQYKTEIDPSMLQKTP  
GNAERKALQQLKAVFAEDNRWVKVAGKGVSETTTGVHRYQWMEKGLLFPAL  
NUNDSYTKSFENLYGCEBSLADGIKRATDVMIAGKVYVVLGYGVGKGCASHMSYSG  
ARVITEDPICALQANEGEVTTWEEAVKEGNIFVTATGNMDVITLDHIKOMRDEA  
IVNIGHFDNEIQTLADNLNFKGATRNIRKQVDRKVFENGNCIYLLAERGLVNLGAT  
GHPFSFYMSNFTNQTLAQILNMQNDKVKVYRLPKKLDDEVARLHLQIGAKLITLTK  
EQADYIGVPEGVKPFHYR"  
complement (3409..4539)  
/gene="metK"  
/note="synonym: CT0722"  
complement (3409..4539)  
/gene="metK"  
/notes="identified by match to PFAM protein family HMM  
PF00438"  
/codon\_start=1  
/transl\_table=11  
/product="S-adenosylmethionine synthetase"  
/protein\_id="AAW1959.1"

```
/db_xref="GI:21646663"
/transl_table=11
/translation="MLDEFIKQDPSNRVACETFTVTTQGVIVGSEVTSKGIIVDVQTIAR
KTIETIGTYEYMFANQSGILSALHSQSDINRGVDRKEBIADBFVRGAGQGM
FYACTTPELPAALQYAOAELVRLLEIKGKIMTYLRPAKSOVTLVDGNDVIL
RVEAVGVSOTQDPEPAGMSEAFQAVIKNDVIENVIKVIKPAKIDENTKEHINPTGR
PEIGGPDGTLGTRKLIIVDTYGGAAHGGGARGKDPKVDKRSAAIYARHVAKNIVA
AGLAKCTQVQSAVAGVAPISYIYNTHGSKGLSDEQIQEKAIAIFDLRPLAIRR
FNLDRPGWCYRDTAAYGHFGREQPFMEKTEKVAELKAALGI"
gene
CDS
complement (4529..4798)
/gene="CT0723"
/gene="CT0723"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAW71960.1"
/db_xref="GI:21646664"
/translation="MDTPKALLRELSPELAERFPYSRLSLRDILHLEKASANRTP
AVLVVRITKRNKGYVTFKVLVYLIGISFRPSGGGKSDFRRRAG"
gene
CDS
4959..5051
/gene="CT0724"
/gene="CT0724"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAW71961.1"
/db_xref="GI:21646665"
/translation="MWMDGVNGLRSEVMRVFCGKTLDPDNNHNL"
gene
CDS
5048..6763
/gene="CT0725"
/gene="CT0725"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAW71962.1"
/db_xref="GI:21646666"
/translation="MNNIGLPRKTVGVVFALVLLSRPSVLWAEAMKFGVNGDTQWT
TADPGCNHTVPSVIRQVNFQIDAGVKRFVIOGDLSDDGKEISBEERVAAPLI
DAGIGFAFNGEAKSAENGYGAPGRQYQPNRGGFTKSDGSFTVGSNPSVPK
ISRLDGLSVSFPFGQGRARFVIIDNPLGRVANSHTYPSGTIADQQVIFAQ
LQKNRKSVTVLSQPLIGHQDILFSGFANEHPWQNRPFESLQSGNVRIFCQ
HDHQRKSVTSPDGSKVEQLIVOSNSKFFYFKSLDITNWFQKREHSVSQERES
VGYIYITDGSFTVDYWDHGHQWSDANFPQAGRADTGYTQFHFVKERWSYL
NGKQPLVPOGASTYTVVRDRFNGEARILDLGNTSEGRDASLDSGQRPVLYVNTGW
IAAPSRHSKGNPAIPVOLFSLGSGSDHTDHLVLSMSYDPAIVROKNIASGFGIV
TQDAAGRWLNVAENGGAATIDGPRWRCVAGLSGIDRKHRAWVILNHEGAPVA
NPSAH"
gene
CDS
6849..7361
/gene="CT0726"
/gene="CT0726"
/notes="identified by match to PFAM protein family HMM
PF01844"
/codon_start=1
/transl_table=11
/product="HNH endonuclease family protein"
/protein_id="AAW71963.1"
/db_xref="GI:21646667"
/translation="MLLOKSKVLVNASYBLSICDARNAVLLFCGKAMVASHPEH
RIRVTENFELPSIVLAWYRIDYRGVLRNLNLPFRDGFQFCYCGCKDGSUTVDHV
MPKRGEDTENWLTACKSNRTKGNRTPSEAGMANKRCPNSNITLAKRHYRSIS
DEWXPYLFMS"
gene
CDS
complement (7405..8382)
/gene="CT0727"
/gene="CT0727"
/notes="identified by match to PFAM protein family HMM
PF00132"
```

```
/codon_start=1
/transl_table=11
/product="glucose-1-phosphate thymidyltransferase,
putative"
/protein_id="AAW71964.1"
/db_xref="GI:21646668"
/translation="MKALIPVAGVGRSLRFTFSQPKVLLNVAAGKPIIGHIMDKLIES
GIDBAVIVGYLGGKIEBYSITLVKVEVDDPRFVVTTGGDRIVLVRKEQPVSNIA
ILGDTIPDVLKLVLGSSITLVKVEVDDPRFVVTTGGDRIVLVRKEQPVSNIA
IVGLYELFRAGTLFNSIDYIITNDIPKGFOLTDALOHMIDLGLPESTPVGWYDC
GKSPRTLLSTNEVLLORDTPROKSLPFCIINPPVFIADRSATVNSIIGNMAIAHAVR
DSLIMNSIIGKRSQVSEIMLDRSLVGNNAIVSAMGHELNIGDYSEIRNG"
gene
CDS
complement (8479..8607)
/gene="CT0728"
/gene="CT0728"
/notes="identified by Glimmer2; putative"
/codon_start=1
Query Match 38.3%; Score 560; DB 1; Length 12968;
Best Local Similarity 63.9%; Pred. No. 2.2e-120;
Matches 914; Conservative 0; Mismatches 480; Indels 36; Gaps 3;
Qy 32 TCGACTACAGAAATTCGCGACATCAACTCCATGTTCTCGGCGCTAAGAACTTACCCCTG 91
Db 3303 TCGACTCAAGGTGGCCGATATTCGCTTCCGAAATGGGGCGCAAGAGATCAAAATCG 3244
Qy 92 CTGAGAGGAAATGCCAGGCTCTTATGGTTCTTCTGAGGGTTATTCGCTTCTTAAGCCAT 151
Db 3243 CTGAAAAGAGATCGCGGCTGATGCTACGCGCAAGAAATACGAAGGCAAGAGCCAC 3184
Qy 152 TGAAGGTGTCAAGATCTCTGTTCCCTCCACATGACGTCCAGACAGCCGCTCTCATCG 211
Db 3183 TGGCAGCGCGCCGATCGCTGCTCCCTGCACATGACCATCCAGACTGCCCTCCTCATCG 3124
Qy 212 AGACACTCACAGCTCTTGTGCTGATGTCAGATGGGCTTCTCGCAACATCTTCTACAC 271
Db 3123 AAGCGCTCTGCAACTCGCGCGGAGCTGCGCTGGGAGAGCTGTAAACATCTTCTGAGC 3064
Qy 272 AAGTACAGCGGCTGCTGCTATGTTGTCGGCCCAACAGGCGACACAGAGAACGAGCG 331
Db 3063 AGGATCAGCGCGCAGCGGCTATCGCTG-----CGCGCG 3031
Qy 332 GTATCCAGCTCTTCGCTCGAAGGGCGAAACATCCCAAGAACTATGGAGAACACATACC 391
Db 3030 GTGTGCCCGGTTCGCTCGAGGGCGAGACCCCTCCAGAGTACTGCTGTCACCCGCTC 2971
Qy 392 GCGCTCTCACATGGCCAGATGTCAAGGCCCAAGCAGAGCTTCTCATGATGCTGGTGTG 451
Db 2970 AGATCTTGAGTTTGAGGGCGCTTCGCGCGGACCTCATCTCATGACGCGCGGACG 2911
Qy 452 CTACACTCTCATCTCCAAGGCTTCGAATTCGAAACAGCCGCTGCTGTTCCAGAGCAA 511
Db 2910 CAACCCCTGATGATCCACTTCGGCTACAAGATCGAAGATGATCTTCGATGCTCGACAAA 2851
Qy 512 CAGAGAGCTGACACCTCGAATACCGCTCGCTTCTTCTTACATCTCAGCAGCTCTTCAAC 571
Db 2850 C---CCCGGCAACGCCGCAAGAAAGCGCTTCCTCCAGCAGCTCAAGGCTCTTTCGCG 2794
Qy 572 AAGCAAGAACCACTGCGCACACAGTTCGCTCGCGGCATGAACGCTGTTTCCGAAGACAA 631
Db 2793 AAGCAACACCGAGCTGCGCACAGGTTGCGCGCGCATGAAGGCTGTCTCCGAAGACCA 2734
Qy 632 CAAAGGTGTCCACCGCTCTACAGGTCCAGAGGAGGCGCAAACTCTCTTCCAGACCA 691
Db 2733 CCACCGCGTGCATCGCTCTACAGATGATGAAAGGGCGAACTGCTTTCCCGGCA 2674
Qy 692 TCAAGCTCAACAGCGCTGTACAAAGTCCCAAGTTCGATACATCTACGGTTCGCGCAT 751
Db 2673 TCAAGCTCAACAGCTCGTCCAGTTCGAGTTCGACAACTTACGGCTTCGAGAGT 2614
Qy 752 CCCTTATCGATGATCAACCGCTGCTCCGATGTCATGATCGGCGCAAGACAGCTCTCG 811
```

```
Db 2613 CGTCCGCTGAGCGATCAAGCGCGCCACGCGACGTGATGATCGCGCGCAAGGTGGTCGTTG 2554
QY 812 TCATGGGTTACGGCGATGTCGGCAAGGGCTGGCTCAATCCCTCCGTGGCCAAGCGGCTC 871
Db 2553 TGGTTGGCTACGGCGAGCTGGCAAGGGCTGGCCCACTCGATGGCTCCTACGGTGGCTC 2494
QY 872 GCGTTATCATCAAGAACTGACCCAACTCTGGCTCTCCAGGCTGCCATGGAAGGCTACC 931
Db 2493 GCGTGTATCGTCAACCGAGATCGACCCGATCTGGCGATTGCAAGCGCGCATGGAGGGTTTG 2434
QY 932 AGTTCGCGCGCATCGAGGAGTGTCAAGGATGTCGATATCTTCGTTAGATGCACAGGAA 991
Db 2433 AGTTCACCACTGGAAGAGCTGTCAAGGAGGCAACAUCTTCGTCACGCCACCGGTA 2374
QY 992 ACTGCGATATCATCTCTGTTGACATGATGGCCAGATGAAGGATPAAGGCTATTGTCGGA 1051
Db 2373 ACAAGGATGTATCAACCTCGACCATCAAGCAGATCGCGCGAGAGGATGTCTGCA 2314
QY 1052 ACATCGGCCACTTCGATTAAGAAATTGATACAGATGGCTCATGAATACCCAGGATCA 1111
Db 2313 ACATCGGCCACTTCGATTAAGAAATTGATACAGATGGCTCATGAATACCCAGGATCA 2254
QY 1112 AGCACATCCCAATCAAGCCAGATACGACATGTGGGAATCCCAAGATGGCCACGCTATCC 1171
Db 2253 CCAGATCAACATCAAGCCGAGGTGACAGTATGTCTTCGAAACGGCACTGCATTT 2194
QY 1172 TCCTTCTTGTAGGGCGGCTTCTTAACTTGGCTGGCTACAGGTCAACCATCTTTCG 1231
Db 2193 ACCTGTCTCGAAGGCGGCTGTGAACTCGGCTGGCCACTGGGCCACCCGTCGTTTCG 2134
QY 1232 TTATGTCAATGTCAATCAAAACAGACACTCGCTCAGTCACTCTACGAAAGAGAG 1291
Db 2133 TGATGAGCACTCTTCCACACAGACGCTGCCCGAGATCGAACTCTGCGAGACGACT 2074
QY 1292 GAAATCTCGAAGAGAGGTTTACACTTCGGAAGCATCTCGATGAAGAGTTCGCTCGCC 1351
Db 2073 ACAGAGTTG-----GTGTCATGCTCGCGAGAGGTTGACGAAGGTTTGGCGGC 2020
QY 1352 TCCACCTCGGATCTCTGATGTCCACCTTCAAGGTTACACAGAAGCGGCTGACTACA 1411
Db 2019 TGCACCTCGGCGAGATTGGCGCCAAAGCTGACACGCTCACCAAGAGCGGCGGACTACA 1960
QY 1412 TCAAGGTTCCAGTTGAGGGTCTTACAAAGTCTGATGCTTACCGTTATTAA 1461
Db 1959 TCGGTGTACGGTTCAGAGGCCCTTCAAGCCCGAGACACTACGCTACTGA 1910
```

Search completed: October 17, 2003, 21:58:12  
Job time : 5479 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: October 17, 2003, 18:46:21 ; Search time 444 Seconds  
(without alignments)

8882.610 Million cell updates/sec

Title: US-09-759-990a-1

Perfect score: 1461

Sequence: 1 atggcttgcaaatcaactac.....ctgatgtaccgttattaa 1461

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Lasting first 45 summaries

Database :

N Geneseq 19Jun03:\*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1459.4	99.9	1461	22	AAS09278 Trichomonas vagina
2	1440.2	98.6	1461	22	AAS09277 Trichomonas vagina
3	596.4	40.8	1557	22	AAF71862 Corynebacterium gl
4	596.4	40.8	1557	23	AAS96121 C. glutamicum gene
5	596.4	40.8	1939	24	ABK52394 sahH gene encoding
6	596.4	40.8	349980	22	AAH68526 C glutamicum codin
7	594	40.7	1422	22	AAH65801 C glutamicum codin
8	558.4	38.2	135638	25	ABX34289 S. atroolivaceus 1

9	535.4	36.6	1488	21	AAA39567 M. tuberculosis an
10	535.4	36.6	4403765	22	AAI99683 Mycobacterium tube
11	535.4	36.6	4411529	22	AAI99682 Mycobacterium tube
12	529.6	36.2	1396	22	AAF71864 Corynebacterium gl
13	529.6	36.2	1396	23	AAS96123 C. glutamicum gene
14	508.8	34.8	1865	21	AAC4037 Zea mays DNA fragm
15	506.6	34.7	1727	21	AAC46315 Arabidopsis thalia
16	497.8	34.1	1728	21	AAC39836 Arabidopsis thalia
17	489	33.5	1767	17	AAT44513 Asparagus S-adenos
18	488.4	33.4	1812	13	AAQ29419 Nicotiana tabacum
19	488.4	33.4	1812	17	AAQ29194 S-adenosylhomocyst
20	471.4	22.3	1494	21	AAC46817 Arabidopsis thalia
21	331	22.7	2200	21	AAC77823 Human cancer assoc
22	331	22.7	2429	25	ABX63344 Human cDNA #344 di
23	331	22.7	2658	23	ABV23126 Human prostate exp
24	331	22.7	2658	23	ABV28967 Human prostate exp
25	317.8	21.8	1302	24	AAD41391 Human NZMS-4 cDNA
26	308.8	21.1	1592	23	ABL28157 Drosophila melanog
27	308.8	21.1	4380	23	ABL28156 Drosophila melanog
28	302	20.7	3830	24	ABS66327 Soybean locus (SyS
29	302	20.7	513445	22	AAI61373 Soybean 318013 reg
30	297.8	20.4	1941	17	AAT44515 Arabidopsis S-aden
31	293.4	20.1	1398	21	AAZ98330 A. thaliana gene i
32	288.4	19.7	720	22	AAH65803 C glutamicum codin
33	278.6	19.1	708	22	AAH65802 C glutamicum codin
34	244	16.7	2583	19	AAV28617 Nucleotide sequenc
35	244	16.7	3616	20	AAV73924 Human SAHH DNA #1
36	244	16.7	3634	23	ABV23195 Human prostate exp
37	244	16.7	3634	23	ABV29032 Drosophila melanog
38	233.6	16.0	1566	23	ABL03693 Drosophila melanog
39	224.2	15.3	1776	23	ABL29787 Human S-adenosylho
40	223.2	15.3	2211	22	AAC91228 Novel human cDNA s
41	213.2	14.6	1976	25	ABX71108 cDNA encoding nove
42	211.6	14.5	1891	24	ABK91312 Human NZMS-14 cDNA
43	189.4	13.0	1188	24	AAD41401 Arabidopsis thalia
44	180.2	12.3	835	24	ABN98911 DNA encoding novel
45	179.2	12.3	5030	23	AAS76216

#### ALIGNMENTS

##### RESULT 1

AAS09278

ID AAS09278 standard; DNA; 1461 BP.

XX AAS09278;

AC AAS09278;

XX 24-OCT-2001 (first entry)

XX Trichomonas vaginalis recombinant His-SAHH gene sequence.

XX SAHH; S-adenosyl homocysteine; S-adenosyl homocysteine hydrolase;

XX S-adenosylmethionine; SAM; glycine N-methyltransferase; GMT; His-SAHH;

XX cancer; malaria; arthritis; SAH; mutant; ds.

XX Trichomonas vaginalis.

OS Synthetic.

XX Key

FT mutation Location/Qualifiers

FT mutation /\*tag= a

FT mutation /\*tag= b

FT mutation /\*tag= c

FT mutation /\*tag= d

FT mutation /\*tag= e

FT mutation /\*tag= f

FT mutation replace (19,G)

FT mutation replace (201,G)

FT mutation /\*tag= b

FT mutation /\*tag= c

FT mutation /\*tag= d

FT mutation /\*tag= e

FT mutation /\*tag= f

FT mutation replace (207,T)

FT mutation /\*tag= c

FT mutation /\*tag= d

FT mutation /\*tag= e

FT mutation /\*tag= f

FT mutation replace (744,T)

FT mutation /\*tag= f

FT mutation replace (834,G)



```

FT mutation /*tag= g
FT replace (897,T)
FT /*tag= h
FT mutation replace (917,T)
FT /*tag= i
FT mutation replace (1206,T)
FT /*tag= j
FT mutation replace (1304,T)
FT /*tag= k
FT mutation replace (1346,T)
FT /*tag= l
XX
PN WO200151651-A2.
XX
PD 19-JUL-2001.
XX
PP 12-JAN-2001; 2001WO-US01114.
XX
PP 14-JAN-2000; 2000US-0176444.
XX
PA (ANTI-) ANTICANCER INC.
XX
XX Hoffman RM, Xu M, Han Q;
XX
XX WPI; 2001-451863/48.
XX
XX Assessing therapeutic levels of S-adenosylmethionine comprises
PT measuring reaction products in sample containing glycine
PT N-methyltransferase, (His) S-adenosyl homocysteine hydrolase and
PT glycine -
XX
PS Claim 6; Fig 6A-6B; 28pp; English.
XX
XX The present sequence for His-SAHH represents a novel recombinant gene
CC sequence for Trichomonas vaginalis S-adenosyl homocysteine
CC (SAHH, S-adenosyl homocysteine hydrolase). The wild type SAHH gene
CC sequence (AA03277) is used to generate the His-SAHH gene
CC which comprises various point mutations and an additional polynucleotide
CC sequence which encodes for an extra 6 N-terminal histidine residues.
CC The invention relates to a method for assessing therapeutic
CC levels of S-adenosylmethionine (SAM) in a biological fluid sample
CC comprising measuring one or more reaction products in a sample
CC containing glycine N-methyltransferase (GMT), SAHH or His-SAHH, and
CC glycine, where the level of one or more products is directly
CC proportional to the level of SAM in the sample. The method is useful
CC for assaying therapeutic levels of SAM in a biological sample and may
CC be used as a part of a diagnostic protocol or as part of a therapeutic
CC protocol, where conditions or progress of the therapy may be monitored.
CC SAHH or His-SAHH may be used as a reagent, particularly in screening for
CC inhibitors and inactivators of the enzyme for use as reagents themselves
CC as potential therapeutics, e.g. in cancer, malaria, arthritis and other
CC diseases. Recombinant SAHH may be used as a therapeutic cancer gene in
CC combination with S-adenosyl homocysteine (SAH) analogues.
XX
SQ Sequence 1461 BP; 372 A; 427 C; 335 G; 327 T; 0 other;
Query Match 99.9%; Score 1459.4; DB 22; Length 1461;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1460; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGGCTTGGAAATCACTACTGTGGTCCATTTCGAGTACAGATTGCGGACATCAACCTC 60
DB 1 ATGGCTTGGAAATCACTACTGTGGTCCATTTCGAGTACAGATTGCGGACATCAACCTC 60
QY 61 CATCTCTCGGCGGTAAAGAACTTACCTCTGCTGAGAGGAAATGCCAGGTCTTATGGTT 120
DB 61 CATCTCTCGGCGGTAAAGAACTTACCTCTGCTGAGAGGAAATGCCAGGTCTTATGGTT 120
QY 121 CTTCGTGAGCGTTATTCCGGTCTTAAGCCATTGAAGGGGTGTCAGAAATCTCTGGTCCCTC 180
DB 121 CTTCGTGAGCGTTATTCCGGTCTTAAGCCATTGAAGGGGTGTCAGAAATCTCTGGTCCCTC 180
QY 181 CACATGACAGTCAGAGAGCGGTCTCTCATCGAGACACTCATCAGGTCTTGGTGTGATGTC 240

```

Db 1261 CTGCTCAGCTCGACTCTACGAAAGAGAGGAAATCTCGAAGAGGTTTACACACTT 1320  
 QY 1321 CCGAAGCATCTCGATGAGAAGTGGCTCGCTCCACCTCGGATCTCTCGATCTCCACCTT 1380  
 Db 1321 CCGAAGCATCTCGATGAGAAGTGGCTCGCTCCACCTCGGATCTCTCGATCTCCACCTT 1380  
 QY 1381 ACAAGCTTACACAGACAGGCTGACTACATCAACGTTCCAGTTGAGGCTCTTACAAAG 1440  
 Db 1381 ACAAGCTTACACAGACAGGCTGACTACATCAACGTTCCAGTTGAGGCTCTTACAAAG 1440  
 QY 1441 TCTGATGCTTACCGTTATTA 1461  
 Db 1441 TCTGATGCTTACCGTTATTA 1461

## RESULT 2

AA09277  
 ID AA09277 standard; DNA; 1461 BP.  
 XX  
 AC AA09277;  
 XX  
 DT 24-OCT-2001 (first entry)  
 XX  
 DE Trichomonas vaginalis wild type SAHH gene sequence.  
 XX  
 KW SAHH; S-adenosyl homocysteinase; S-adenosyl homocysteine hydrolase;  
 KW S-adenosylmethionine; SAM; Glycine N-methyltransferase; GMT; His-SAHH;  
 XX cancer; malaria; arthritis; SAH; ds.  
 OS Trichomonas vaginalis.

## TH Key Location/Qualifiers

FT mutation replace (19,A)  
 FT /tag= a  
 FT mutation replace (201,C)  
 FT /tag= b  
 FT mutation replace (207,C)  
 FT /tag= c  
 FT mutation replace (210,C)  
 FT /tag= d  
 FT mutation replace (501,T)  
 FT /tag= e  
 FT mutation replace (744,C)  
 FT /tag= f  
 FT mutation replace (834,C)  
 FT /tag= g  
 FT mutation replace (897,A)  
 FT /tag= h  
 FT mutation replace (917,C)  
 FT /tag= i  
 FT mutation replace (1206,C)  
 FT /tag= j  
 FT mutation replace (1304,A)  
 FT /tag= k  
 FT mutation replace (1346,C)  
 FT /tag= l

WO200151651-A2.

EN  
 XX PD 19-JUL-2001.

XX PF 12-JAN-2001; 2001WO-US01114.

XX PR 14-JAN-2000; 2000US-0176444.

XX PA (ANTI-) ANTICANCER INC.

XX PI Hoffman RM, Xu M, Han Q;

XX DR WPI; 2001-451863/48.

XX PT Assessing therapeutic levels of S-adenosylmethionine comprises  
 FT measuring reaction products in sample containing glycine

PT N-methyltransferase, (His) S-adenosyl homocysteine hydrolase and  
 PT glycine  
 XX Disclosure, fig 6A-6B; 28pp; English.  
 PS  
 XX The present sequence represents the gene sequence for Trichomonas  
 CC vaginalis wild type S-adenosyl homocysteinase (SAHH, S-adenosyl  
 CC homocysteine hydrolase). The wild type SAHH gene sequence is  
 CC used to generate a novel recombinant SAHH gene sequence, His-SAHH,  
 CC (AA09277) which comprises various point mutations and an additional  
 CC polynucleotide sequence which encodes for an extra 6 N-terminal histidine  
 CC residues. The invention relates to a method for assessing therapeutic  
 CC levels of S-adenosylmethionine (SAM) in a biological fluid sample  
 CC comprising measuring one or more reaction products in a sample  
 CC containing glycine N-methyltransferase (GMT), SAHH or His-SAHH, and  
 CC glycine, where the level of one or more products is directly  
 CC proportional to the level of SAM in the sample. The method is useful  
 CC for assaying therapeutic levels of SAM in a biological sample and may  
 CC be used as a part of a diagnostic protocol or as part of a therapeutic  
 CC protocol, where conditions or progress of the therapy may be monitored.  
 CC SAHH or His-SAHH may be used as a reagent, particularly in screening for  
 CC inhibitors and inactivators of the enzyme for use as reagents themselves  
 CC as potential therapeutics, e.g. in cancer, malaria, arthritis and other  
 CC diseases. Recombinant SAHH may be used as a therapeutic cancer gene in  
 CC combination with S-adenosyl homocysteine (SAH) analogues.  
 XX  
 SQ Sequence 1461 BP; 369 A; 420 C; 338 G; 334 T; 0 other;

Query Match 98.6%; Score 1440.2; DB 22; Length 1461;  
 Best Local Similarity 99.1%; Pred. No. 0;  
 Matches 1448; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 ATGGCTTGCAAAATCACCTACTGGTCTCCATTCGAGTACAGAAATTCGCGATCAACCTC 60  
 Db 1 ATGGCTTGCAAAATCACCTACTGGTCTCCATTCGAGTACAGAAATTCGCGATCAACCTC 60  
 QY 61 CATGTTCTCGGCGGTAAAGAACTTACCTTGTCTGAGAGGAATGCCAGGCTTATGGTT 120  
 Db 61 CATGTTCTCGGCGGTAAAGAACTTACCTTGTCTGAGAGGAATGCCAGGCTTATGGTT 120  
 QY 121 CTTGTGAGGGTTATTCCGCTTTCTAAGCCATTGAAGGGTGTGAGAATCTCTGGTTCCTC 180  
 Db 121 CTTGTGAGGGTTATTCCGCTTTCTAAGCCATTGAAGGGTGTGAGAATCTCTGGTTCCTC 180  
 QY 181 CACATGACAGTCCAGACACCGCTCTCTATCGAGACACTCAGAGCTCTTGGTGTGATGTC 240  
 Db 181 CACATGACAGTCCAGACACCGCTCTCTATCGAGACACTCAGAGCTCTTGGTGTGATGTC 240  
 QY 241 AGATGGGCTTCCTGCAACATCTTCTACACAAGATACAGCGCTGTGCTGTATCGTTGTC 300  
 Db 241 AGATGGGCTTCCTGCAACATCTTCTACACAAGATACAGCGCTGTGCTGTATCGTTGTC 300  
 QY 301 GGCCCAACAGGCAACACAGAGAAGCCAGCGGTATCCAGTCTTCGCTTGAAGGGCGAA 360  
 Db 301 GGCCCAACAGGCAACACAGAGAAGCCAGCGGTATCCAGTCTTCGCTTGAAGGGCGAA 360  
 QY 361 ACACCTCCAGANTACTGGGAGACACATACCGCGCTCTCACATGCCAGATGTCAGGC 420  
 Db 361 ACACCTCCAGANTACTGGGAGACACATACCGCGCTCTCACATGCCAGATGTCAGGC 420  
 QY 421 CCACAGCAGGTTTCTGATGATGGTGTGATGCTTACACTCTCTCATCTCCAGGGCTTCGAA 480  
 Db 421 CCACAGCAGGTTTCTGATGATGGTGTGATGCTTACACTCTCTCATCTCCAGGGCTTCGAA 480  
 QY 481 TTCGAAACAGCCGGTCTGTTCCAGAGCCCAACAGAGCTTGACACCTCGAATACCGGTGC 540  
 Db 481 TTCGAAACAGCCGGTCTGTTCCAGAGCCCAACAGAGCTTGACACCTCGAATACCGGTGC 540  
 QY 541 GTTCTTGTACTACTCAAGCAGGTCTTCAACCAAGACCAAGAACCTGTCACACAGTTGCT 600  
 Db 541 GTTCTTGTACTACTCAAGCAGGTCTTCAACCAAGACCAAGAACCTGTCACACAGTTGCT 600  
 QY 601 GCCGGCATGAACGGTGTGTTTCCGAGAGAGACAAACAGGTGTCCACCGCTCTACAGCTC 660

```
Db 601 GCGCGCATGAACGGTGTTCGAGAGAGACAACACAGGTGTCCACCGCTCTACACAGTCC 660
Qy 661 GAGAAGGGGCAAACTCTCTTCCAGCCATCAAGCTCAAGCGCTGTTCACAAAGTCC 720
Db 661 GAGAAGGGGCAAACTCTCTTCCAGCCATCAAGCTCAAGCGCTGTTCACAAAGTCC 720
Qy 721 AAGTTCGATAACATCTCGGCTCGCCACTCCCTTATCGATGTATCAACCGTCTTCC 780
Db 721 AAGTTCGATAACATCTCGGCTCGCCACTCCCTTATCGATGTATCAACCGTCTTCC 780
Qy 781 GATGTCAATGATCGGGCAAGACAGACTCTCGTCAATGCTTACGGCGATGTCGGCAAGGCG 840
Db 781 GATGTCAATGATCGGGCAAGACAGACTCTCGTCAATGCTTACGGCGATGTCGGCAAGGCG 840
Qy 841 TGGGCTCAATCCCTCCGTCGCGGCAAGGCTCGGCTTATCATCAAGAACTCCAGCCCAATC 900
Db 841 TGGGCTCAATCCCTCCGTCGCGGCAAGGCTCGGCTTATCATCAAGAACTCCAGCCCAATC 900
Qy 901 TGGGCTCTCCAGGCTGCCATGGAAGGCTACAGGCTCCGCGCATCGAGGAAGTTCGTCAG 960
Db 901 TGGGCTCTCCAGGCTGTCTATGGAAGGCTACAGGCTCCGCGCATCGAGGAAGTTCGTCAG 960
Qy 961 GATGTGATATCTTCGTTACATGACAGGAAGTCCGATATCATCTCTGTCACATGATG 1020
Db 961 GATGTGATATCTTCGTTACATGACAGGAAGTCCGATATCATCTCTGTCACATGATG 1020
Qy 1021 GCCCAGATGAAGGATTAAGCTTATGTCGTTAATCATCGGCCACTTCGATACAGGAATTCAT 1080
Db 1021 GCCCAGATGAAGGATTAAGCTTATGTCGTTAATCATCGGCCACTTCGATACAGGAATTCAT 1080
Qy 1081 ACAGATGGCTCATGAATACCGAGGATCAAGGATCAAGCATCCCAATCAAGCCAGAAATACGAC 1140
Db 1081 ACAGATGGCTCATGAATACCGAGGATCAAGGATCAAGCATCCCAATCAAGCCAGAAATACGAC 1140
Qy 1141 ATGTGGGAATCCAGATGGCCAGCTATCTCTTCTGCTGAGGCGCGCTTCTTAAAC 1200
Db 1141 ATGTGGGAATCCAGATGGCCAGCTATCTCTTCTGCTGAGGCGCGCTTCTTAAAC 1200
Qy 1201 CTTGGCTGGCTACAGGTCACCCATCTTCTGTTATGTCATGTCATTCACAAACAGACACA 1260
Db 1201 CTTGGCTGGCTACAGGTCACCCATCTTCTGTTATGTCATGTCATTCACAAACAGACACA 1260
Qy 1261 CTGCTCAGTCACTCTAGCAAGAGAGAGAAATCTCGAGAGAGAGTTTACACATTT 1320
Db 1261 CTGCTCAGTCACTCTAGCAAGAGAGAGAAATCTCGAGAGAGAGTTTACACATTT 1320
Qy 1321 CCGAAGCATCTCGATGAAGAAAGTCGCTCGCTCCACCTCGGATCTCTCGATGTCCACCTT 1380
Db 1321 CCGAAGCATCTCGATGAAGAAAGTCGCTCGCTCCACCTCGGATCTCTCGATGTCCACCTT 1380
Qy 1381 ACAAGGTTACAGAGAGAGGCTGACTACATCAAGCTTCAGTGGGTCCTTACAG 1440
Db 1381 ACAAGGTTACAGAGAGAGGCTGACTACATCAAGCTTCAGTGGGTCCTTACAG 1440
Qy 1441 TCTGATGCTTACCGTTATTAA 1461
Db 1441 TCTGATGCTTACCGTTATTAA 1461
```

## RESULT 3

AAAF1862

ID AAF1862 standard; DNA; 1557 BP.

XX AAF1862;

AC AAF1862;

XX 30-APR-2001 (first entry)

DT 30-APR-2001 (first entry)

XX Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:219.

DE Corynebacterium glutamicum; metabolic pathway protein; MP protein;

KW fine chemical production; microorganism; organic acid; nucleoside;

KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;

KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;  
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.  
XX Corynebacterium glutamicum.  
XX WO200100843-A2.  
XX 04-JAN-2001.  
XX 23-JUN-2000; 2000WO-IB00923.  
XX 25-JUN-1999; 99US-0141031.  
PR 01-JUL-1999; 99DE-1030476.  
PR 02-JUL-1999; 99US-0142101.  
PR 08-JUL-1999; 99DE-1031415.  
PR 08-JUL-1999; 99DE-1031418.  
PR 08-JUL-1999; 99DE-1031419.  
PR 08-JUL-1999; 99DE-1031420.  
PR 08-JUL-1999; 99DE-1031424.  
PR 08-JUL-1999; 99DE-1031428.  
PR 08-JUL-1999; 99DE-1031434.  
PR 08-JUL-1999; 99DE-1031435.  
PR 08-JUL-1999; 99DE-1031443.  
PR 08-JUL-1999; 99DE-1031453.  
PR 08-JUL-1999; 99DE-1031457.  
PR 08-JUL-1999; 99DE-1031465.  
PR 08-JUL-1999; 99DE-1031478.  
PR 08-JUL-1999; 99DE-1031510.  
PR 08-JUL-1999; 99DE-1031541.  
PR 08-JUL-1999; 99DE-1031573.  
PR 08-JUL-1999; 99DE-1031592.  
PR 08-JUL-1999; 99DE-1031632.  
PR 08-JUL-1999; 99DE-1031634.  
PR 08-JUL-1999; 99DE-1031636.  
PR 09-JUL-1999; 99DE-1032125.  
PR 09-JUL-1999; 99DE-1032126.  
PR 09-JUL-1999; 99DE-1032130.  
PR 09-JUL-1999; 99DE-1032186.  
PR 09-JUL-1999; 99DE-1032206.  
PR 09-JUL-1999; 99DE-1032227.  
PR 09-JUL-1999; 99DE-1032228.  
PR 09-JUL-1999; 99DE-1032229.  
PR 09-JUL-1999; 99DE-1032310.  
PR 14-JUL-1999; 99DE-1032922.  
PR 14-JUL-1999; 99DE-1032926.  
PR 14-JUL-1999; 99DE-1032928.  
PR 14-JUL-1999; 99DE-1033004.  
PR 14-JUL-1999; 99DE-1033005.  
PR 14-JUL-1999; 99DE-1033006.  
PR 12-AUG-1999; 99US-0148613.  
PR 27-AUG-1999; 99DE-1040764.  
PR 27-AUG-1999; 99DE-1040765.  
PR 27-AUG-1999; 99DE-1040766.  
PR 27-AUG-1999; 99DE-1040832.  
PR 31-AUG-1999; 99DE-1041378.  
PR 31-AUG-1999; 99DE-1041379.  
PR 31-AUG-1999; 99DE-1041380.  
PR 31-AUG-1999; 99DE-1041394.  
PR 31-AUG-1999; 99DE-1041396.  
PR 03-SEP-1999; 99DE-1042076.  
PR 03-SEP-1999; 99DE-1042077.  
PR 03-SEP-1999; 99DE-1042079.  
PR 03-SEP-1999; 99DE-1042086.  
PR 03-SEP-1999; 99DE-1042087.  
PR 03-SEP-1999; 99DE-1042088.  
PR 03-SEP-1999; 99DE-1042095.  
PR 03-SEP-1999; 99DE-1042124.  
PR 03-SEP-1999; 99DE-1042129.  
PR 09-MAR-2000; 2000US-0187970.  
XX (BADI ) BASF AG.  
PA Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
XX



XX PA (BAD1) BASF AG.

XX PI Pompeius M, Kroege B, Schroeder H, Zelder O, Haberhauer G, Kim J;

XX PI Lee H, Hwang B;

XX DR WPI; 2001-582269/65.

XX DR P-PSDB; AAU71911.

XX PT Nucleic acids encoding metabolic pathway proteins from *Corynebacterium*

XX PT glutamicum, useful for producing methionine and lysine in

XX PT *Corynebacterium* and *Brevibacterium*.

XX PS Disclosure; Page 276-278; 316pp; English.

XX CC The present invention relates to the isolation of novel *Corynebacterium*

XX CC glutamicum genes encoding metabolic pathway (MP) proteins

XX CC (AAU71863-AAU71922). The metabolic pathway proteins of the invention

XX CC include enzymes involved in the lysine and methionine biosynthetic

XX CC pathways. The polynucleotide sequences of the invention can be used

XX CC for the large-scale production and/or modulation of expression of

XX CC fine chemicals such as lysine and methionine. The sequences of the

XX CC invention may be used to identify *C. glutamicum* and related organisms

XX CC e.g. *C. diphtheriae* in a subject to detect diphtheria.

XX CC AA96073-AA96132 represent *C. glutamicum* genes encoding the novel

XX CC metabolic pathway proteins of the invention.

XX SQ Sequence 1557 BP; 348 A; 440 C; 431 G; 338 T; 0 other;

Query Match 40.8%; Score 596.4; DB 23; Length 1557;

Best Local Similarity 64.5%; Pred. No. 1.3e-165;

Matches 923; Conservative 0; Mismatches 501; Indels 6; Gaps 2;

QY 32 TCGAGTACAGAAATCCGACATCAACTCCATGTTCTTCGCGGTAAGAACTTACCCCTG 91

DB 114 TGGACTTCAAGTTCCGATCTTCACTAGCAGAGCAGGACGTCACCAATTCGTCCTG 173

QY 92 CTGAGAGGAATGCCAGGCTTATGTTCTTCGTCGAGCTTATTCGCTTCAGACCAT 151

DB 174 CAGAGTATGAGATGCCAGGCTCTCATGCAATTCGCGAGAAATTCGACAGAGACGCTT 233

QY 152 TGAAGGGTCTCAGAAATCTCTGGTTCCTCCATGACAGTCCAGACAGCGCTCTCATCG 211

DB 234 TGAAGGGGCGCCGAATTCGTTCTATCCATGACGCTCCAGACCGCGCTGCTTATG 293

QY 212 AGACATCTCAGCTCTTGGTCTGATCTCAGATGGCTTCTTCCAGATCTTCTPACAC 271

DB 294 AGACCTCTCACTGTTTGGCGCTGAGGTTCTGTTGGGCTTCTTCCAAATTTCTCCACC 353

QY 272 AAGATACAGCGCTGCTGATCTGTTGTCGCGCCCAACAGGACACACAGAGCCAGCGG 331

DB 354 AGGATAGGCTGACAGCGCTATCTGTTCTGGC---TCGCGACCTCGAGAGCCAGCTG 410

QY 332 GTATCCAGTCTTGGCTTGAAGGGCGAAACATCTCCAGATACTGCGGAAACATATCC 391

DB 411 GTGTTCCAGTATTCGCGTGAAGGCTGAGTCACTGAGGAGTACTGGTGTGTCATCAAC 470

QY 392 GCGCTCTCAGTGGCCAGATGTTCAAGGCCCAACAGAGCTTCTCGATGATGTTGTTGATG 451

DB 471 AGATCTTACGCTGGGCGATGAGC---TGCCAAACATGATCTCGACAGCGGCTGAGC 527

QY 452 CTACATCTCTCATCTCCAGGGCTTCGAATTCGAATTCGAATTCGAATTCGAATTCGA 511

DB 528 CCACCATGGCTGTTATTCGCGTCCGAAATACGAGCAGGCTGTTGTTGTTCCACAGCAG 587

QY 512 CAGAGCTGACACCTCGAATACCGCTGCTTCTTCTTCTACTCAGCAGGCTTCAACC 571

DB 588 AGGCCAACGATTCGCGATGATGATCGCATTTCTTGGGCATGCTCGTGAAGTCTTGTG 647

QY 572 AAGCAAGAACCACCTGACACAGTTCGTCGCGCATGAACGCTGTTTCCGAGAGACAA 631

DB 648 CAGAGCTTGCAGATGGGCGAGATCGCTGAGGCGCTTAAAGGCTGTCACCGAGAAACA 707

QY 632 CAACAGGTGTCCACCGCTCTACAGCTCGAGAGGAGGGCAAACTCCTCTTCCAGCCA 691

DB 708 CCACCGGTGTGCACCGCTGTACCACTTCGCTGAAGAGGCTGCTGCTTCCAGCCA 767

QY 692 TCACGCTCAACGAGCGCTGTACAAAGTCCAAAGTCCATTAATCTACGGCTGCGCCACT 751

DB 768 TGAACGTCAACGAGCGCTGTACCAAGTCCAAAGTTCATTAACAGTACGCGACCGCCACT 827

QY 752 CCCTTATCGATGTATCAACCGTGTCCGATGTCTATGATCGCGGCAAGACAGCTCTCG 811

DB 828 CCCTGATCGAGGCATCAACCGCGCACTGACATGCTCATGGCGCGAGAGCGTGTG 887

QY 812 TCATGGGTTCAGCGGATGTGCGCAAGGCTGCGCTCAATCCCTCCGTTGGCCAGAGCGCTC 871

DB 888 TCTCGGTTACGCGGATGTGCGCAAGGCTGCGCTGAGGCTTTCGACGCGCAGCGGCTC 947

QY 872 GGGTTCATCATCAGAACTCGAACCAATCTCGCGCTCTCCAGGCTGCCATGAGGGCTACC 931

DB 948 GCGTCAAGGTACCGAAGCTGACCAATCAACGCTTTCAGGCTCTGATGATGGCTACT 1007

QY 932 AGGTCCGCGCATCGAGGAAGTCTGTCAGGATGTCATATCTTCGTTACATGACACAGAA 991

DB 1008 CTGTGCTCACGTTGATGAGGCAATCGAGGAGCGGACATCGTATCACCGCACCGCA 1067

QY 992 ACTCGATATCATCTCTGTTGACATGATGCGCCAGATGAGGATGAGCTTATGTCGTA 1051

DB 1068 ACAAGGACATCATTTCTTCGAGCAGATGCTCAAGATGAAGGATCACGCTCTGCTGGCA 1127

QY 1052 ACATCGGCACTTCGATAACGAATTGATACAGATGGCTCATGAAATACCAGGACATCA 1111

DB 1128 ACACTGCTCACTTGTATATGATGATGATGATGATGATGATGATGATGATGATGATG 1187

QY 1112 AGCATCTCCATCAAGACGAGATGAGCATGAGGAAATTCAGATGAGGCGACGCTATCC 1171

DB 1188 CCGCACACGATCAAGACGAGGTCGACGAGTTCACTTCTCCACCGGCTGCTCCATCA 1247

QY 1172 TCTTCTTGTGAGGGCGGCTTCTTAACCTTGGCTGCGCTACAGTCAAGCTCACTTCTTCG 1231

DB 1248 TGTCTCTGTCGAGGTCGCTTGTGAACCTTGGCAAGCCGACCGGACACCCATCAITG 1307

QY 1232 TTATGTCATGTCATTCACAAACGAGACCTGCTGCTGAGTTCGACCTTACGAAAGAGAG 1291

DB 1308 TCATGTCACACTTTTCGCGGATCAGACCATTCGCGAGATCGAACTGTTTCCAAACGAG 1367

QY 1292 GAAATCTCGAGAGAGGTTTACACACTTCGAGGATCTCGATGAGAGTTCGCTCGCC 1351

DB 1368 GACATGACAGAGGAGGCTACCGCTCTGCTTAAGGTTCTGACGAAAGGTGGACCCA 1427

QY 1352 TCCACTCGGATCTCTCGATGTCCACCTTACAAAGCTTACACAGAGCAGGCTGACTACA 1411

DB 1428 TCCAGTTTGAAGGCTCTCGCGGTCAGCTCACCGAATGACCAAGGAGCAGGCTGAGTACA 1487

QY 1412 TCAAGTTTCAAGTTGAGGCTCTTCAAGTCTGATGCTTACCGTTATTA 1461

DB 1488 TCGGCTGTGAGTGTGACGCGCCCAATTAAGCGGAGACACTACCGTACTAA 1537

RESULT 5

ABK52394

ID ABK52394 standard; DNA; 1939 BP.

XX AC ABK52394;

XX AC ABK52394;

XX DT 27-AUG-2002 (first entry)

XX DE SdhH gene encoding adenosyl homocysteine enzyme.

XX KW SdhH; adenosyl homocysteine; L-amino acids production; L-lysine;

XX KW L-methionine; animal feedstuff; fermentation; gene; ds.

XX OS *Corynebacterium glutamicum*.

XX FH Key Location/Qualifiers

FT	CDS	227..1723	
FT		/*tag= a	
FT		/product= "Adenosyl homocysteinease"	
FT		/FT	
XX		/transl_except= {pos:227..229, aa:Met}	
XX			
PN		WO200220806-A1.	
PD			
FD		14-MAR-2002.	
XX			
PF		17-JUL-2001; 2001WO-EP08222.	
PR			
PR		09-SEP-2000; 2000DE-1044706.	
PR		28-FEB-2001; 2001DE-1009685.	
XX			
PA		(DEGS ) DEGUS5A AG.	
XX			
PI		Farwick M, Huthmacher K, Brehme J, Pfefferle W, Binder M;	
PI		Greissinger D, Thierbach G;	
DR		WPI: 2002-463087/49.	
DR		P-PDSB; AAU97842.	
XX			
PT		Novel sahH gene from coryneform bacteria useful as probe to isolate	
PT		genes coding for adenosyl homocysteinease, and overexpression of which	
PT		gene in coryneform bacteria is useful for producing amino acids, e.g.	
PT		L-lysine	
XX			
PS		Claim 4; Page 45-47; 53pp; English.	
XX			
CC		The invention describes an isolated polynucleotide (I) from coryneform	
CC		bacteria, comprising a sequence coding for the sahH gene. A coryneform	
CC		bacteria in which the sahH gene is enhanced (II) is useful for preparing	
CC		L-amino acids, in particular L-lysine and L-methionine for e.g.	
CC		L-methionine-containing animal feedstuffs. The method involves fermenting	
CC		(II), concentrating L-amino acids in the medium or in the cells of the	
CC		bacteria, and isolating the L-amino acid. In this method, bacteria in	
CC		which further genes of the biosynthesis pathway of the desired L-amino	
CC		acid are additionally enhanced, or in which the metabolic pathways which	
CC		reduce the formation of the desired L-amino acid are partly eliminated	
CC		are employed. The bacteria is transformed with a plasmid vector which	
CC		carries the nucleotide sequence coding for the sahH gene. Corynebacterium	
CC		produces amino acids in an improved manner after over-expression of the	
CC		sahH gene. This sequence represents the Corynebacterium glutamicum sahH	
CC		gene that encodes the enzyme adenosyl homocysteinease.	
XX			
XX		Sequence 1939 BP; 428 A; 559 C; 530 G; 422 T; 0 other;	
QY			
QY		Query Match 40.8%; Score 596.4; DB 24; Length 1939;	
QY		Best Local Similarity 64.5%; Pred. No. 1.5e-165;	
QY		Matches 923; Conservative 0; Mismatches 501; Indels 6; Gaps 2	
QY			
QY		32 TCAGGTACAGAAATTGGCGACATCAACTCCATGTTTCGGCGGTAAGAACTTACCCCTG 91	
Db		300 TGGACTTCAGGTTGGCGATCTTTCACTAGCAGAGGACGAGTCACAGATTGTCGTTG 359	
QY		92 CTCAGAGGAAATCCGAGGCTTATGGTCTTCGTGAGCGGTATTCCGCTTCTAAGCAT 151	
Db		360 CAGAGTATGAGATCCGAGGTCTCATCGAGTTGGCGAAGAAATCCGAGACGAGAGCCTT 419	
QY		152 TGAAGGGGTGCAGAAATCTCTGGTTCCTCCACATGACAGTCCACAGACGCGTCCTCATCG 211	
Db		420 TGAAGGGCGCCGAAATGCTGGTCTTATCCACATGACGGTCCAGACCGCGTCTATTG 479	
QY		212 AGACATCAGAGCTTTGGTGTGATGATGATGAGTGGGCTTCTCGAACATCTTCTACAC 271	
Db		480 AGACCTCCTACCTGCTTTGGCGGCTGAGGTGCTGTTGGGCTTCTCTGCAATTTCTCCACC 539	
QY		272 AAGATACAGCCGCTGCTGCTATGCTTGTTCGCGCCACACAGGCACACGAGAGCCAGCG 331	
Db		540 AGATGAGGCTTGACAGGCTATGCTGTTGTGCG---TCGCGCACGCTCGAAGAGCCAGCTG 596	
QY		332 GTATCCCATGCTCTTCGCTTGAAGGGCGAAACACTCCCGAATATCTGGGAGAACACATACC 391	



Db 189927 GACAGTACGAGAACAGGCTCTACCGTCTGCTAAGGTTCTCGACGAAAGGTGGCAGCA 189986  
QY 1352 TCCACCTCGGATCTTCGATGTCACCTTACAAAGCTTACACAGAACGAGGTGACTACA 1411  
Db 189987 TCCACGTTGAGGCTCTCGGCGTTCAGTCCAGCACTGACCAAGGAGGAGGCTGAGTACA 190046  
QY 1412 TCAACGTTCCAGTTCGAGGTCCTTACAAAGTCTGATGCTTACCGTTATTA 1461  
Db 190047 TCGGCTTTCAGCTTCAGGCCCCATTCAAGCGGAGCAGTACCGTACTTAA 190096

RESULT 7  
AAH65801  
ID AAH65801 standard; DNA; 1422 BP.  
XX AAH65801;  
XX 26-SEP-2001 (first entry)  
XX C glutamicum coding sequence fragment SEQ ID NO: 836.  
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis; ds.  
XX Corynebacterium glutamicum.  
XX EP1108790-A2.  
XX 20-JUN-2001.  
XX 18-DEC-2000; 2000EP-0127688.  
XX 16-DEC-1999; 99JP-0377484.  
PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX WPI; 2001-376931/40.  
DR P-PSDB; AAG90582.  
XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
PT  
XX Claim 8; SEQ ID NO: 836; 246pp + Sequence Listing; English.  
XX The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived  
CC from coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
XX Sequence 1422 BP; 309 A; 414 C; 401 G; 298 T; 0 other;  
SQ

Query Match 40.7%; Score 594; DB 22; Length 1422;  
Best Local Similarity 64.5%; Pred. No. 6.5e-165;  
Matches 920; Conservative 0; Mismatches 500; Indels 6; Gaps 2;  
QY 32 TCGAGTACAGATTGCCGACATCAACCTCCAGTCTCGGCGGTAAGGAACTTACCCTTG 91

Db 2 TGGACTTCAAGGTTGCCGATCTTTCACTAGCAGCAGCAGCTCACAGATTGCTTG 61  
QY 92 CTGAGAGGAAATGCCAGGCTTATGGTTCTTGGTGGAGGTTATTCCGCTTCTTAGCCAT 151  
Db 62 CAGAGTATGAGATGCCAGGCTCTCATGTCAGTTGGCGAAGAAATTCGACAGCAGCAGCTT 121  
QY 152 TGAAGGGTGTGAGATCTCTGGTTCCCTCCACATGACAGTCCAGACAGCCCTCTCATCG 211  
Db 122 TGAAGGGCGCCGAAATGCTGCTTATCCACATGACGCTCCAGACCGCGCTGTTATG 181  
QY 212 AGACATCAACAGCTCTTGGTCTGATGTCAGATGGGCTTCCTGGAAACATCTTCTACAC 271  
Db 182 AGACCTCACTGCTTTGGCGCTGAGGTTGTTGGGCTTCTGCAACATTTCTCCACC 241  
QY 272 AAGATACAGCGGCTGCTGCTATCTGTTGGGCCCAACAGSCACACAGAGAACGACGCG 331  
Db 242 AGGATGAGGCTGCAGCGGCTATCGTTTCGGC--TCCGGCACCGTCGAGAGCCAGCTG 298  
QY 332 GTATCCAGTCTTGGCTCGAAGGGCGAAACACTTCCCAAGATATCTGGGAGAACATACC 391  
Db 299 GTGTTCCAGTATTCGGCTGGAAGGGTGAATCACTGAGGAGTACTGGTGGTGCATCAAC 358  
QY 392 GCGCTCTCAATGCGCAGATGTCAGGCCCAACAGCAGGTTGTCGATGATGGTGGTATG 451  
Db 359 AGATCTTCAGCTGGGCGATGAGCT--GCCAACATGATCTCTGACACGCGGCGTACG 415  
QY 452 CTACACTCTCATCTCCAAGGGCTTTCGAATTCGAAACAGCGGCTCTTCCAGAGCCAA 511  
Db 416 CCACCATGGCTGTTATTTCCGGCTCGCAATACGAGCAGGCTGTTCTGTTCCACAGCAG 475  
QY 512 CAGAAGCTGACAACTTCGAATACCGCTCGCTTCTTGTACTACTCAAGCAGGCTTTCACAC 571  
Db 476 AGGCCAAAGATTCGATGAGTATCATCGCATTTCTGGGCATGCTCGTGAGGTTCTTGCTG 535  
QY 572 AAGACAAGAACCACTGCGCACACAGTTGCTCGCGCATGAACGGTGTTCGGAAGAGCAA 631  
Db 536 CAGAGCTGGCAAGTGGGCGAGATCGCTGAGGCCGTTAAGGGTGTACCGAGGAACCA 595  
QY 632 CAACAGGTGTCACCGCTTCTACAGCTTCGAGAGAGAGGGGCAAACTCTCTTCCAGCCA 691  
Db 596 CCACCGGTGTGCAACCGCTCTACCACTTCGCTGAAGAGGCGTGTGCTTCCAGCGCA 655  
QY 692 TCACGCTCAGCAGCGCTTACAAAGTCCAGTTCGATTAACATCTACGGCTGCGGCACT 751  
Db 656 TGAACGTCACAGCGCTGTCACCAAGTCCAAAGTTGATTAACAGTACCGCACCCGCACT 715  
QY 752 CCTTATCGATGGTATCAACCGTGTCTCCGATGTCATGATCGGCGGCAAGACAGCTCTCG 811  
Db 716 CCTGATCGACGGCATCAACCGCGCACTGACATGCTCATGCGCGGCAAGAAAGTGTCTTG 775  
QY 812 TCATGGGTACGGGATGTCGGCAGGGCTGCGCTCAATCCCTCGTGGCCAGCGGCTC 871  
Db 776 TCTGGGTTACGGGATGTCGGCAAGGGCTGCGCTGAGGCTTTCGACGGCGAGGGGCTC 835  
QY 872 GCGTTATCATCACAGAACTCGAACCAATCTCGCTCTCCAGGCTGCCATGGAAGGTTACC 931  
Db 836 GGTCAAGGTACAGAGCTGACCCCAATCAACGCTCTTCAGGCTCTGATGATGGTACT 895  
QY 932 AGGTCCGCGCATCGAGGAGTGTCAAGGATGTCGATATCTTCGTTACATGACAGGAA 991  
Db 896 CTGTGTCACCGTTGATGAGGCGCATCGAGGACCGCGACATCGTATCACCGCAGCGCA 955  
QY 992 ACTCGGATATCATCTCTGTTGACATGATGGCCAGATGAAGATGAAGGCTATTTGTGGTA 1051  
Db 956 ACAAGCATCATTTCTCTCGAGCAGATGCTCAAGATGAAGATCACGCTCTGCTGGCA 1015  
QY 1052 ACATCGGCCACTTCGATAACGAAATTTGATACAGATGGCTCATGAAATACCCAGGATCA 1111  
Db 1016 ACATCGGTCACTTGTGATATGAGATCGATATGATTTCCCTGTTGACCGCGACGACGTCA 1075  
QY 1112 AGCATCCCAATCAAGCCAGATACGACATGTGGGAATTTCCAGATGGCCAGCTATCC 1171



Db 1076 CCGCACACAGATCAGGCCACAGGTCGACGAGTTCACCTTCTCCACGGTCTCCATCA 1135  
 Qy 1172 TCCTTCTGCTGAGGGCGGCTTCTTAACTTGGCTGGCTACAGGTTCACCGATCTTTG 1231  
 Db 1136 TCGTCTCTGCGAAGTGGCTGTGTAACCTTGGCAACGCCACCGACACCCATCAITG 1195  
 Qy 1232 TTATGTCAATGTCATACAAACAGACACTGCTCAGCTCAGCTTACGAAAGAGAG 1291  
 Db 1196 TCATGTCCAACTCTTTCGCGCATCAGCAATTCGCGCATGCAACTGTTCGAAACGGA 1255  
 Qy 1292 GAAATCTCCAGAGAGGTTTACACACTTCCGAACTCTCGATGAAGAGTCTGCTGCC 1351  
 Db 1256 GACAGTACAGAACAGAGTCTACCGTCTGCTTAAAGTTCTCGACGAAAGTGGACCCA 1315  
 Qy 1352 TCCACCTCGGATCTCTCGATTCACCTTACAAAGCTTTACACGAGAGGCTGACTACA 1411  
 Db 1316 TCCACGTTGAGGCTCTGGCGGTTCAGCTCACCGAACTGACCAAGGAGCGGTGAGTACA 1375  
 Qy 1412 TCAAGCTTCCAGTTGAGGTCCTTACAGTCTGATGCTTACCGTTA 1457  
 Db 1376 TCGCGCTTGAAGTTGAGGCGCCATTCAGCGCGGAGCACTACCGCTA 1421

RESULT 8  
 ABX34289/c  
 ID ABX34289 standard; DNA; 135638 BP.  
 XX AC ABX34289;  
 XX DT 11-FEB-2003 (first entry)  
 XX DE S. atroolivaceus leinamycin biosynthesis gene cluster.  
 XX KW Leinamycin biosynthesis gene cluster; lmm; open reading frame; ORF;  
 KW anti-tumour antibiotic; broad spectrum antimicrobial activity;  
 KW Gram-positive; Gram-negative bacteria; chemical modification;  
 KW metabolite; apo-carrier protein; holo-carrier protein; tumour;  
 KW polyketide; hybrid polypeptide/polyketide metabolite; lmm production;  
 KW cytosolic; gene; ds.  
 XX OS Streptomyces atroolivaceus.  
 XX PN WO200277179-A2.  
 XX PD 03-OCT-2002.  
 XX PF 22-MAR-2002; 2002WO-US08937.  
 XX PR 26-MAR-2001; 2001US-278935P.  
 XX PA (REGC ) UNIV CALIFORNIA.  
 XX PA (KYOW ) KYOWA HAKKO KOGYO KK.  
 XX PI Shen B, Cheng Y, Tang G;  
 XX WPI; 2003-018907/01.  
 DR P-PSDB; ABU11341, ABU11342, ABU11343, ABU11344, ABU11345, ABU11346.  
 DR ABU11347, ABU11348, ABU11349, ABU11350, ABU11351, ABU11352, ABU11353,  
 DR ABU11354, ABU11355, ABU11356, ABU11357, ABU11358, ABU11359, ABU11360,  
 DR ABU11361, ABU11362, ABU11363, ABU11364, ABU11365, ABU11366, ABU11367,  
 DR ABU11368, ABU11369, ABU11370, ABU11371, ABU11372, ABU11373, ABU11374,  
 DR ABU11375, ABU11376, ABU11377, ABU11378, ABU11379, ABU11380, ABU11381,  
 DR ABU11382, ABU11383, ABU11384, ABU11385, ABU11386, ABU11387, ABU11388,  
 DR ABU11389, ABU11390, ABU11391, ABU11392, ABU11393, ABU11394, ABU11395,  
 DR ABU11396, ABU11397, ABU11398, ABU11399, ABU11400, ABU11401, ABU11402,  
 DR ABU11403, ABU11404, ABU11405, ABU11406, ABU11407, ABU11408, ABU11409,  
 DR ABU11410, ABU11411.

XX Novel gene cluster responsible for synthesis of leinamycin in  
 PT Streptomyces atroolivaceus useful for making various peptide and/or  
 PT polyketide, and/or hybrid polypeptide/polyketide metabolites  
 XX Claim 6; Page 81-127; 185pp; English.

XX The present invention relates to the isolation of the Streptomyces  
 CC atroolivaceus leinamycin (lmm) biosynthesis gene cluster containing  
 CC 71 open reading frames (ORFs) (ORFs -35 through -1, ORFs lmmA through  
 CC lmmZ, and ORFs +1 through +9). Leinamycin is a novel anti-tumour  
 CC antibiotic produced by several Streptomyces species. It exhibits  
 CC broad spectrum antimicrobial activity against Gram-positive and  
 CC Gram-negative bacteria, but not against fungi. The polypeptides encoded  
 CC by the lmm biosynthesis gene cluster ORFs are useful for chemically  
 CC modifying a molecule in a host cell. The host cell is a bacterium or  
 CC eukaryotic cell, including a mammalian, yeast, plant, fungal, or insect  
 CC cell. The molecule is an endogenous metabolite produced by the host  
 CC cell or exogenously supplied metabolite, or an amino acid, and the  
 CC polypeptide is a peptide synthetase or amino transferase. The  
 CC polypeptides encoded by the lmm gene cluster are useful for converting  
 CC an apo-carrier protein to a holo-carrier protein. lmm shows potent  
 CC antitumour activity in tumour models in vivo. The lmm gene cluster  
 CC modules and/or catalytic domains are useful for making various peptide  
 CC and/or polyketide, and/or hybrid polypeptide/polyketide metabolites.  
 CC The proteins encoded by the ORFs are useful alone, or in combination  
 CC with other active domains to modify various target substrates. The  
 CC lmm gene cluster is useful to upregulate endogenous lmm production to  
 CC permit lmm production in cells and/or to make various modified lmm.  
 CC lmm, its analogue, or other polyketide, peptide or hybrid  
 CC polyketide/peptide metabolites are useful as therapeutic agents, to  
 CC treat a number of disorders, depending upon the type of metabolites.  
 CC The present sequence represents the S. atroolivaceus leinamycin  
 CC biosynthesis gene cluster.  
 XX SQ Sequence 135638 BP; 18570 A; 49096 C; 49039 G; 18933 T; 0 other;  
 Query Match 38.2%; Score 558.4; DB 25; Length 135638;  
 Best Local Similarity 63.7%; Pred. No. 2.6e-153;  
 Matches 909; Conservative 0; Mismatches 471; Indels 48; Gaps 2;  
 Qy 34 GAGTACAGAAATGCGACATCAACCTCCATGTTCTCGGCGCTAAGGAACCTTACCTTGCT 93  
 Db 131697 GACTTCAGGTCGCGGACCTGTCCCTCCGCGCTTCGGCGCAGAGATCACCTCGCC 131638  
 Qy 94 GAGAAGGAATGCGAGGTCTTATGGTTCTTGTGAGGCTTATTCGCTTTAAACCAATTG 153  
 Db 131637 GAGCAGGAGATGCGCGGCTGATGTCGATCCGCGAGAGTAGCGCGCGCGCACCCGCTG 131578  
 Qy 154 AAGGTGTGAGATCTCTGTTCCCTCCACATGATCCAGACAGCGCTCTCATCGAG 213  
 Db 131577 GCGGTGCCAGGATCACCGGTCCTTCGATGATGATGATGATGATGATGATGATGATG 131518  
 Qy 214 ACATTCACAGCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 273  
 Db 131517 ACCCTGTCGCGCTTCGCGCGGACGTCGCTGGTCTCTCTGCAACATCTACTCCACCCAG 131458  
 Qy 274 GATACAGCGCGCTGCTGATTCGTTGTCGGCCCAACAGGCACACAGAGAGCCAGCGGT 333  
 Db 131457 GACCAAGCGCGCGCGGATGCGCGG-----CCGCGGGC 131425  
 Qy 334 ATCCAGCTCTTCGCTCGAAGGCGGAAACACTCCAGAACTACTGGGAGAACACATACCGC 393  
 Db 131424 ATCCCGGTCTTCGCTTGAAGGCGGAGCCTCGAGAGGTACTGTGTGTCACGAGCAG 131365  
 Qy 394 GTCTCAATGCGCAGATGTTCAAGGCCCAACAGAGGTGTTGATGATGTTGTTGTTGTTGTTG 453  
 Db 131364 GGGCTGACTTGGCGCGGCCACACCGCGCGGAAACATGATCTCTGAGACGCGCGGAGTGC 131305  
 Qy 454 ACATCTCTTCTCTCCAGGSGCTTCGAATTCGAAACAGCGGTCTGTTCCAGAGCCAAACA 513  
 Db 131304 ACCCTCTCTGTCACAGAGGTGTCGAGTACCGCAAGCGGGATGATGCGCGGCGGAG 131245  
 Qy 514 GAAGCTGACAACTTCGAATACCGCTCGCTTCTTCTTACTCAAGCAGGTCTTCAACCAA 573  
 Db 131244 AAGAGGAACTGGCGGTGTTGTCGCGCC-----CTGTGGACCGC 131206  
 Qy 574 GAACAAGAACCTCGCACACAGTTGTCGCGGATGAAACGGTGTTCGCGAGAGACAAACA 633

Db 131205 AGCGGACTCGATGCGAGCGCCATGTCTCTCGGAGATCCCGGTGTGACGGAGGAGACCAAG 131146  
 QY 634 ACAGGTGTCCACCGCTCTACAGCTCGAGAAGAGGAGGCAAACTCTCTCTTCCAGCCATC 693  
 Db 131145 ACCGGTGTCCACCGCTCTACAGAGTGACCGGAGCGACCCCTGTCTGTTCGGCGATC 131086  
 QY 694 AACGTCAAGACGCTGTGTACAAAGTCAAGTTCGATTAACATCTACGGCTGCGCCCACTCC 753  
 Db 131085 AATGTGAAGACGCGCTGACGAAGTCAAGTTCGATTAACATCTACGGCTGCGCCCACTCC 131026  
 QY 754 CTTATCGATGATCAACCGCTGTCTCGATGTCTCGATGTCTCGATGTCTCGATGTCTCGAT 813  
 Db 131025 CTATCGAGGGATCAACCGCTGTCTCGATGTCTCGATGTCTCGATGTCTCGATGTCTCGAT 130966  
 QY 814 ATGGGTGTAGCGGATGTCTCGAGGCTGTCTCGATGTCTCGATGTCTCGATGTCTCGATGTCT 873  
 Db 130965 TGGCGGTACGGGAGTGGGCAAGGGGTGTCTCGATGTCTCGATGTCTCGATGTCTCGATGTCT 130906  
 QY 874 GTTATCATCAGAACTGACCCATCTCGGCTCTCGAGGTCTCGAGGTCTCGAGGTCTCGAGGT 933  
 Db 130905 GTGATCATCAGGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 130846  
 QY 934 GTCGCGGATCATGAGGAGTGTCTCAAGGATGTCTCGATTCGATTCGATTCGATTCGATTCGAT 993  
 Db 130845 GTGGGACGCTCGAGAGTGTCTCGAGGAGTGTCTCGATTCGATTCGATTCGATTCGATTCGAT 130786  
 QY 994 TGGGATATCATCTCTGTGTGATGATGCGCCGAGATGAAGGATTAAGGCTATTTCGGTAAAC 1053  
 Db 130785 AAGGACATCATGTGCGCTCGGATGCGGCAAGATGAAGCACCAGGCGATGTGGGCAAC 130726  
 QY 1054 ATGGGCACTTCGATTAAGAAATGATACAGATGCGCTCTATGAATACCAAGGATCAAG 1113  
 Db 130725 ATCGGCACTTCGATTAAGAAATGATACAGATGCGCTCTATGAATACCAAGGATCAAG 130666  
 QY 1114 CACATCCCAATCAAGCAGATGATGAGATGTGGGAATTCACAGATGCGCAAGGCTATCTCTC 1173  
 Db 130665 AAGGACGAGTCAAGCGCAGGTCCACCTGGAAGTTCGCCGATGCGAAGGTTCCTCATC 130606  
 QY 1174 CTTCTGTGAGGGCGCTTCTTAACCTTGGCTGGCTACAGTCAACCACTTTCTGTT 1233  
 Db 130605 GTCTGTGAGGGCGCTTCTTAACCTTGGCTGGCTACAGTCAACCACTTTCTGTT 130546  
 QY 1234 ATGTCAATGATTCACAAACACGACACTGCTGAGTCTGAGTCTTACGAAAGAGAGA 1293  
 Db 130545 ATGTCAATGATTCACAAACACGACACTGCTGAGTCTTACGAAAGAGAGA 130486  
 QY 1294 AATCTCGAAGAAGGTTTACACACTTCCGAAGCTCTCGATGAAGAGTTCGCTGCGCTC 1353  
 Db 130485 GAGTACCGGACCGAGTCTAGTGTCTTCCCAAGCACTTCGACGAGAGGTTCGCGCTG 130426  
 QY 1354 CACCTCGGATCTCTGATGTCCACCTTACAAAGCTTACAGAGGAGGCTGACTACATC 1413  
 Db 130425 CACCTCGGATCTCTGATGTCCACCTTACAAAGCTTACAGAGGAGGCTGACTACATC 130366  
 QY 1414 AAGCTTCCAGTGTAGGGTCTTACAAAGTCTGATGTTTACCGTTATTAA 1461  
 Db 130365 GGGGTAGAGTTCGAGGCGCGGTACAGCGGACCACTACCGTACTGA 130318

RESULT 9  
 AAA39567  
 ID AAA39567 standard; DNA; 1488 BP.  
 XX  
 AC AAA39567;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE M. tuberculosis antigen TB54 DNA.  
 XX  
 DE Tuberculosis; TB; antigen; vaccine; diagnosis; somatic; tuberculostatic;  
 KW infection; interferon-gamma; IFN-gamma; protective immunity; therapy;  
 XX delayed type hypersensitivity response; TB54; ds.  
 XX

OS Mycobacterium tuberculosis.  
 XX WO200021983-A2.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 08-OCT-1999; 99WO-DK00538.  
 XX  
 PR 08-OCT-1998; 98DK-0001281.  
 XX  
 PR 21-JAN-1999; 99US-0116673.  
 XX  
 XX (STAT-) STATENS SERUM INST.  
 XX  
 PI Andersen P, Weidling K, Hansen CV, Florio W, Okkeis LMM;  
 PI Skjot RL, Rosenkrands I;  
 XX P-PSDB; AA87873.  
 DR WPI: 2000-317931/27.  
 XX  
 XX Novel polypeptide of somatic protein extract useful as vaccine against  
 PT virulent Mycobacterium infection, isolated from cell wall, cell  
 PT membrane and cytosol -  
 XX  
 PS Claim 6; Page 94-96; 126pp; English.  
 XX  
 CC This invention describes a novel polypeptide (PP) of somatic proteins  
 CC extract (I) which have tuberculostatic activity. (I) or their subsequence  
 CC has at least one of the following properties: (a) the PP induces an in  
 CC vitro recall response, or an in vitro response, during primary infection  
 CC with virulent Mycobacterium, determined by a release of interferon  
 CC (IFN)-gamma, (b) PP induces a protective immunity, determined by  
 CC vaccinating an animal with PP and an adjuvant, three times at two weeks  
 CC intervals, (c) PP induces an in vitro response, or in vitro recall  
 CC response, determined by release of IFN-gamma of at least 1000 pg/ml or  
 CC 500 pg/ml, respectively, from Peripheral Blood Mononuclear Cells (PBMC)  
 CC withdrawn from TB patients, or PPD positive individuals, 6 months after  
 CC diagnosis, (d) PP induces a specific antibody response in a TB patient,  
 CC as determined by enzyme linked immunosorbent assay (ELISA) technique or  
 CC a western blot, (e) PP induces a positive delayed type hypersensitivity  
 CC (DTH) response, determined by intradermal injection. (I) and (II) are  
 CC useful in preparing a prophylactic or therapeutic vaccine as a vaccine  
 CC for induction of a protective or generation of an immune response in a  
 CC mammal against infection with a virulent Mycobacterium. (I) and (II) are  
 CC also useful as diagnostic reagent for the diagnosis of a virulent  
 CC Mycobacterium infection. The vaccine of the invention induces efficient  
 CC immunological memory, providing long term protection against TB. This  
 CC sequence encodes a Microbacterium tuberculosis TB54 antigen described  
 CC in the invention.  
 XX  
 SQ Sequence 1488 BP; 313 A; 475 C; 474 G; 226 T; 0 other;  
 Query Match 36.8%; Score 535.4; DB 21; Length 1488;  
 Best Local Similarity 61.3%; Pred. No. 1.4e-147;  
 Matches 879; Conservative 0; Mismatches 551; Indels 3; Gaps 1;  
 QY 32 TCGAGTCAGAAATGCGGACATCAACCTCCATGTCTTCGGCCGTGAAGAACTTACCCTTG 91  
 Db 56 TCGACTTTAAGATCGCGGACCTGTCACTAGCGGATTCGCGCGCAAGAACTTCGGATCG 115  
 QY 92 CTGAGAGGAATGCCAGGTCTTATGGTTCTTGTGAGGCTTATTCGGCTTCTAAGCCAT 151  
 Db 116 CCGAGCAGAGATGCCCGGCTGATGTCGTGCGGCGGAGTATCGGAGGTGCAACCCC 175  
 QY 152 TGAAGGGTGTGAGAACTCTGCTTCCCTCCACATGACAGTCCAGACAGCCCTCTCATCG 211  
 Db 176 TGAAGGGGGCCCGGATCTCGGGTTCGCTGCACATGACGCGTGCAGACCGCGGTGTGATCG 235  
 QY 212 AGACACTCACAGCTCTTGGTGTGATGTGATGGGCTTCCTGCAACATCTTCTCTACAC 271  
 Db 236 AAACCTTCACGCGCTGGGCGCGGAGTCCGCTGGGCTGTGCAACATCTTCTTCACCC 295  
 QY 272 AAGATACAGCGCTGTGCTGCTATGCTTGTGCGGCCAACAGGCAACACGAGAGGCGCGCG 331

Db 296 AGGATCACGCGCGCGCGCGCTGCTGCTGGCCCGCGCACCGGCACCCCGACGAGCCCGAAGG 355  
QY 332 GTATCCCACTCTTCGCTCGAAGGCGAACAACCTCCAGATACTCTGGGGAACACATACC 391  
Db 356 GTGTCCCGGTGTTCCGCTGGAGGCGGAGCGCTGAGAGTACTGTGGCGCGCGAGC 415  
QY 392 GCGCTCTCACATGGCCAGA---TGGTCAAGGCCCAACAGAGTTGTCGATGATGGTGTG 448  
Db 416 AGATGCTCACTTGGCGGAGCCCGACAAAGCGGSCAAACATGATCTCGATGAGCGGCTG 475  
QY 449 ATGCTACATCTCTCCAGGCTTCGAATTCGAACAGCGCGTGTCTGTTCCAGAGC 508  
Db 476 ACGCCACCATGTTGTGTGTGCTGCGGCGATGCGATGATGAGAAGCGCGGTGTGCGCCCG 535  
QY 509 CAACAGAGTGCACAACTCGAATACCGCTGCGTCTTCTGCTACACTCAACGAGGTTTCA 568  
Db 536 CCGAGGAGGACGACCCCGCGAGTGAAGTCTTCTGAACCTGCTACGAGCCCGCTTCG 595  
QY 569 ACCAGACAGAACCACTGGCACACAGTTGCTGCGGCAATGAACGTTGTTTCGAGAGA 628  
Db 596 AGACCGACAGGACAAAGTGGACCAAGATAGCGAGTCCGTCGAAGGGGCTCACCGAGGAG 655  
QY 629 CAACACAGGTGTCACCGCTCTAACAGCTCGAGAGGAGGCAAACTCTCTTCCAG 688  
Db 656 CCACCCCGGCTGTCTGGGCTTACCAATTGCGCGCGCGGGATCTGGCTTCCCG 715  
QY 689 CCATCAACGTCAAGACGCTGTTCACAAAGTCCAAAGTTCGATAACATCTACGGTGC 748  
Db 716 CGATCAACGTCAACGCTCGTGGTACCAAGTCCAAATTCGACAAAGTACGGCACTCG 775  
QY 749 ACTCCCTTATCGATGATCAACGCTCTTCGAGTTCATGATCGCGGCGACAGAGCTC 808  
Db 776 ACTCCCTGATCGAGGCTCAACCGCGGACCGCGCTGATCGCGGTGAAGAGTCC 835  
QY 809 TCGTCATGGTTACGCGGATTCGCGAAGGCTGCGCTCAATCCCTCCGTCGCGCAAGCG 868  
Db 836 TCATCTGCGCTACCGGACGCTGCTAAGGCTGTGCGAGCGATCAAGGGCCAGGAG 895  
QY 869 CTCGCTTATCATCAGAACTCGACCCATCTGCGCTCTCCAGCTGCGCATGGAAGCT 928  
Db 896 CGCGGTCTCGCTACCGAGATCGACCCGATCAAGCGCTGACGCGCATGATGGAGGT 955  
QY 929 ACCAGGTCCCGCATCGAGGAAGTCTCAAGGATTCGATCTTCTGTTACATGCACAG 988  
Db 956 TCGAGTGTCTACCGCTCGAGGCGCATCGGAGCGCGACATCGTCTTAACCGCACCG 1015  
QY 989 GAATCTCGATATCATCTCTGTTGACATGATGCGGCGACATGAAGATGAGCTATTGTCG 1048  
Db 1016 GCACAAAGACATCATCTCTCGAGCACATTAAGCGGATGAAGGACCGCGCATCTGG 1075  
QY 1049 GTAACATCGGCCACTTCGATACGAATTCATACATGGCTCATGAATATCCAGGCA 1108  
Db 1076 GAATATCGGCCACTTCGACACAGATCGACATGCGCGGCTGAGCGCTCCGGGCGA 1135  
QY 1109 TCAAGCACATPCCAATCAAGCGAGAAATGACATGTTGGAATTCACAGATGGCCACGCTA 1168  
Db 1136 CACGGGTCAACGTCGAAGCTCAGTTCGACCTGTGACCTTTGGCGACACAGCGCGCTGA 1195  
QY 1169 TCCTCTCTTCTGTGAGGCGCGCTTCTTAACCTTGGCTGCTACAGGTCACCATCTT 1228  
Db 1196 TCATCTGCTGTCCGAGGGGGGCTGTGTGAACCTTGGGCAATCGCACCGGACCCCTCGT 1255  
QY 1229 TCGTTATGTCATATGTCATTCAAAACACAGACACTCGCTCAGCTCGACTCTACGAAAGA 1288  
Db 1256 TCGTGATGAGCAACASCTTCGCTAACACAGACGATCGCCAGATCGAGCTGTGACCAAGA 1315  
QY 1289 GAGGAATCTCGAGAGAGGTTTACACATTCGGAAGCATCTCGAGAGATGAGAGTCCGCTC 1348  
Db 1316 ACACGAGTACGACAAACGAGGTGTACCGGCTGCGCAAGCACCTCGAGAGAGGTTGGCTC 1375  
QY 1349 GCTTCCACCTCGGATCTCTCCGATGTCACCTTACAAAGCTTACACAGACAGCGCTGACT 1408  
Db 1376 GAATCCATGTGAGGGCCCTTGGCGGTCACTGTGACCAAGCTGACCAAGAGCGGCCGAT 1435

QY 1409 ACATCAAGTCCAGTTGAGGTCTCTTACAGTCTGATGCTTACCGTTATTAA 1461  
Db 1436 ACTCGCGTGCAGCTCGAGGTCCTTACAGCGGACCACTACCGCTACTGA 1488

# RESULT 10

AAI99683/c  
ID AAI99683 standard; DNA; 4403765 BP.  
XX AAI99683;  
AC AAI99683;  
XX  
DT 15-JAN-2002 (first entry)  
XX  
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.  
XX  
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
variation; epidemiology; patient treatment; epidemic monitoring; ds.  
XX  
OS Mycobacterium tuberculosis.  
XX  
FN US6294328-B1.  
XX  
PD 25-SEP-2001.  
XX  
PF 24-JUN-1998; 98US-0103840.  
XX  
PR 24-JUN-1998; 98US-0103840.  
XX  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Fleischmann RD, White OR, Fraser CM, Venter JC;  
XX  
DR WPI; 2001-647261/74.  
XX  
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises  
determining the nucleotide sequence of the strain at positions in the  
genome corresponding to positions where M. tuberculosis strains CDC  
1551 and H37Rv differ -  
XX  
PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.  
XX  
CC The invention relates to evaluating strain variation within and between  
different populations of the tuberculosis bacterial pathogen.  
CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
nucleotide sequence of the first strain at positions in the complete  
sequence of the genome that correspond to positions that differ in the  
nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and  
H37Rv (AAI99682). The method is useful for evaluating strain variation of  
M. tuberculosis and has valuable application in the fields of  
tuberculosis genetics, epidemiology, patient treatment and epidemic  
monitoring.  
CC Note: The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from USPTO  
at [seqdata.uspto.gov/sequence.html?DocID=6294328B1](http://seqdata.uspto.gov/sequence.html?DocID=6294328B1).  
XX  
SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

# Query Match

Best Local Similarity 36.6%; Score 535.4; DB 22; Length 4403765;  
Matches 879; Conservative 0; Mismatches 551; Indels 3; Gaps 1;

QY 32 TCGAGTACAGAAATGCGGACATCAACCTCCATGTTCTCGCGCGTAAGAACTTACCCCTTG 91  
Db 3624437 TCGACTTTAAGATCGCGACCTGTCACTAGCGGATTTGCGCGCAAGAACTTCGGATCG 3624378  
QY 92 CTGAGAGGAATAACCAAGTCTTATGTTCTTCTGAGCGTTATTCGCTTCTTAAGCCAT 151  
Db 3624377 CCGAGCAGAGATCCCGGCGCTGATGCTGCGGCGCGAGTATGCCAGGTGCAACCC 3624318  
QY 152 TGAAGGGTGTGAGAAATCTCGTTCCCTCCACATGACAGTCAGACAGCCGCTCATCG 211  
Db 3624317 TGAAGGGGCCCCGAGATCTCGGGTTCGCTGACATGACGGTGCAGACCGCGGTGTGATCG 3624258

```
QY 212 AGCACTCACAGCTCTTGGTCTGATCTCAGATGGGCTTCTCGAATCTTCTCTACAC 271
Db 3624257 AAACCCCTCACCGCGCTGGCGCGCGAAGTCCGCTGGGCTCTGTGACATCTTCTCCACC 3624198
QY 272 AAGATACAGCGCTGCTGCTATCGTTGTGGGCCCAACAGSCACACCCAGAGAGCCAGCG 331
Db 3624197 AGGATCAGCGCGCGCGCGCTGTGTGGCGCGCGCAGCCACCCCGAGAGCCCAAGG 3624138
QY 332 GTATCCAGTCTTCCCTGGAAGGGGAAACACTCCAGATACTCGGAGACACATACC 391
Db 3624137 GTGTCCCGGTGTTCGCGTGAAGGGCGAGACCTCGAAGATGTACTGTGGCGCGCGGAGC 3624078
QY 392 CGCTCTCATATGGCCAGA--TGTCAAGGCCACACAGAGTGTGCGATGATGTGGTG 448
Db 3624077 AGATGCTCACTTGGCGGACCCCGACAGCGCGGCCAACATGATCTCGATGAGCGGCTG 3624018
QY 449 ATGCTACATCTCTCATCTTCCAGGGCTTCGAATTCGAACAGCGGCTGTCTTCAGAGC 508
Db 3624017 ACGCCACCATGTGTGTGTGCGCGGCATGCAGTATGAGAAGGCGCGCGTGTGGCGCGCG 3623958
QY 509 CAACAGAGCTGACAACTCGAATACCGTGTCTTGTGTACACTCAAGCAGGTCTTCA 568
Db 3623957 CCAGAGAGACGACCCCGCGAGTGAAGGTCTTCTGAACCTGCTACGACCCGCTTCG 3623898
QY 569 ACCAAGACAAGAACCACTGGCACACAGTTGCTGCCGCGCATGAACGGTGTTCGAAGAGA 628
Db 3623897 AGACCGACAAAGCAAGTGGACCAAGATAGCGGAGTCTGCTCAAGGGCGTCAACGAGGAGA 3623838
QY 629 CACACACAGTGTCCACCGCTCTTACCAGCTCGAAGAGGGGCAACTCTCTTCCAG 688
Db 3623837 CCACCAACCGGCTGTGTGGGCTTACCAATTCGCGCGCGCGGGGATCTGGGCTTCCCGG 3623778
QY 689 CCATCAAGCTCAACAGACCTCTTACAAAGTCCAAAGTTCGATAAATCTAGGCTCGCGCC 748
Db 3623777 CGATCAACGTCAACAGCTCGGTGACCAAGTTCGAAATTCGACAAAGTACGCGCTCGGC 3623718
QY 749 ACTCCCTTATCGATGATCAACCGTGTCTTCGATGTGATGATCGCGGGCAGACAGCTC 808
Db 3623717 ACTCCCTGATCGACGGCATCAACCGCGCACCGCGCTGATCGCGGCTGAAGAGGTCC 3623658
QY 809 TCCTCATGTGTACGCGCATGTGCGCAAGGGCTGGCTCAATCCTCGTGGCGCAAGCG 868
Db 3623657 TCATCTGCGGCTACGCGCATGTGCTGTAAGGCTGTGCGGAGCGCATGAAGGCGCCAGGAG 3623598
QY 869 CTCGGTATATACAGAACTCGAACCAATCTGCGCTCTCCAGGCTGCGCATGGAAGGCT 928
Db 3623597 CGCGGGTCTCCGTCAACCGAGATCGACCCGATCAACCGCGCTGAGGCGCATGATGAGGGCT 3623538
QY 929 ACCAGTCCGCGCATCGAGAGTGTCTCAGAGTGTGATATCTTGGTTACATGACAG 988
Db 3623537 TCAGTGTGTACCGTTCAGAGAGCCCATCGGGACCGCATCGTGTGTAACCGGACCG 3623478
QY 989 GAAACTGCGATATCTCTCTTGACATGATGCGCCCGATGAAGGATTAAGGCTATTGTG 1048
Db 3623477 GCACAAAGACATCATCTGCTCGAGCACATTAAGCGCATGAAGGACCAACGCGATCTTG 3623418
QY 1049 GTAACATCGGCACCTTCGATTAACCAATTGATACAGATGGCTCATGAATATCCAGGCA 1108
Db 3623417 GAAATATCGGCACCTTCGACAAACAGATCGACATGCGCGGGCTGGAGCGCTCGGCGGCA 3623358
QY 1109 TCAAGCACATCCCAATCAAGCCAGATACGACATGFGGAATTCGCCAGATGCGCACGCTA 1168
Db 3623357 CACGGGTCAACGTCAAGCTCAGTTCGACCTGTGACCTTTGGGCGACACGGCGCGTGA 3623298
QY 1169 TCCTCTCTTGTGAGGCGCGCTTTTAACTTGGTGTGCTGCTACAGGTCAACCATCTT 1228
Db 3623297 TCATGTGTGTTCGAGGGGGGCTGTGTGAACCTGGGCAATGCCACGGGCAACCCCTGT 3623238
QY 1229 TCGTTATCTCAATGCTATTCACAACAGACACTTGCCTCAGCTGACCTCTACGAAGA 1288
Db 3623237 TCGTATCTAGCAACAGCTTCTGCTAAACAGACGATGCCACAGATCGCCACAGATGAGCTGTGAGCAAGA 3623178
```

```
QY 1289 GAGAAATCTCGAGAAGAGGTTTACACACTTCCGAAGCATCTCGATCAAGAGTGGCTC 1348
Db 3623177 ACGACGAGTAGACACACGAGGTGTACCGGTGCGCCAGCACTCGACGAGAGTGGCTC 3623118
QY 1349 GCCTCCACCTCGGATCTTCTGATGTCACCTTACAAAGCTTACAGAGCAGGCTGACT 1408
Db 3623117 GAATCCATGTGAGAGGCCCTTGGCGGTCACTGACCAAGCTGCCAAGGAGCAGGCCGAAT 3623058
QY 1409 ACATCAAGTTCAGTGGGTCCTTACAAAGTCTGATGCTTACCGTTATTAA 1461
Db 3623057 ACCTCGGCTGAGCTCGAGGTCCTTACAGCCGACCACTACCGTACTGA 3623005

RESULT 11
AAI99682/c
ID AAI99682 standard; DNA; 4411529 BP.
XX
AC AAI99682;
XX
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
XX
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-0103840.
XX
PR 24-JUN-1998; 98US-0103840.
XX
PA (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
DR WPI; 2001-647261/74.
XX
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ -
XX
PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;
```

```
Query Match 36.8%; Score 535.4; DB 22; Length 4411529;
Best local similarity 61.3%; Pred. No. 1.le-145;
Matches 879; Conservative 0; Mismatches 551; Indels 3; Gaps 1;

QY 32 TCGAGTCAGAAATGCGCAGATCAACCTCCATGTTCTCGCGCGTAAGAACTTACCCTTG 91
Db 3629588 TCGATTTTAGATCGCGGACCTGTCACTAGCGGATTCGGCGCGCAAGAACTCCGATCG 3629529
```

QY 92 CTGAGAGGAATGCCAGGCTTATGTTCTTCGTGACCGCTTATTCGGCTTCTAGCCAT 151  
Db 3629528 CCGAGCAGGATGCGCGGCTGTGTCGCGCGAGTATGCGGAGGTGCAACCCC 3629469  
QY 152 TGAAGGGTGTCAAGATCTCTGGTTCCCTCCACATGACAGTCCAGACAGCGCTCTCATCG 211  
Db 3629468 TGAAGGGGCCCGATCTCGGTTTCGTGCAATGACGTTGCAACCGGGTGTGATCG 3629409  
QY 212 AGACATCAGAGTCTTGCTGCTGATGTGAGTGGGCTTCCTGCAACATCTTCTACAC 271  
Db 3629408 AAACCTCTCAACCGCTGGGCGCCGAAAGTCCGCTGGGCTCTGTGCAACATCTTCTCCACC 3629349  
QY 272 AAGATACAGCGCTGCTGCTATCGTTCTGGGCCCAACAGGACACACAGAGAACCCAGCG 331  
Db 3629348 AGGATCAGCGCGCGCGCTCTGTGTCGCGCCGACCGCACCCCGAGAGCCCAAGG 3629289  
QY 332 GTATCCAGTCTTGGCTGTGAAGGGCGAAGACACTCCGAGATCTGGGAGAACATACC 391  
Db 3629288 GTGTCCGCTGTTCGCTGGAAGGGCGAGACGCTCGAAGAGTACTGCTGGGCGCGCCGAGC 3629229  
QY 392 GCGCTCTCACATGGCCGAGA--TGGTCAAGGCCACACAGGTTGTCGATGATGTTGGTG 448  
Db 3629228 AGATGCTACCTTGGCGGAGCCCGACAGCCGCGCAACATGATCTCGATGAGCGGGTG 3629169  
QY 449 ATGCTACATCTCTCAATCTTCCAGGGCTTGAATCGAAACAGCGGCTGTCTCCAGAGC 508  
Db 3629168 ACGCCACCATGTTGCTGCGCGCATGCAATGATGAGAGGCGCGGCTGTGCGGCCG 3629109  
QY 509 CAACAGAGCTGACACCTCGAATACCGTGGTCTTCTGACACTCAACAGAGTCTTCA 568  
Db 3629108 CCGAGGAGGACGCCCGCGAGTGAAGGTCTTCTGAACTCTACGACCGCTTCG 3629049  
QY 569 ACCAAGACAAAGAACCACTGGCACAGTGTCTGCGGCAATGAACGCTGTTCGAAAGAGA 628  
Db 3629048 AGACCGCAAGCAAGTGGACCAAGATAGCGAGTCTGCTCAAGGGGCTCAACGAGAGA 3628989  
QY 629 CAACACAGGTTCCACGCTCTACACAGCTCGAAGAGGAGGCAACCTCTTCCAG 688  
Db 3628988 CCACACCGGCTGTGCGGCTCTACCAATTCGCGCGCGGGATCTGGCTTCCCGG 3628929  
QY 689 CCATCAAGTCAACGACGCTGTATCAAGTCCAAAGTTCGATAACATCTACGGTCCGCGC 748  
Db 3628928 CGATCAACGTCACGACCTCGGTGACCAAGTCCAAATTCGACAAAGTACGCGCACTCGGC 3628869  
QY 749 ACTCCCTTATCGATTCGAACCGCTCTTCGATGTCATGATCGGCGGCAACAGAGCTC 808  
Db 3628868 ACTCCCTGATCGACGCGATCAACCGCGCACCGCGCTGATCGGCGGTAAGAGGTC 3628809  
QY 809 TCGTCATGGTTACGCGGATCTCGGCAAGGCTCGGCTCAATCCCTCGGTGGCCAGGCG 868  
Db 3628808 TCATCTCGGCTAGCGGACGCTGCTGTAAGGCTGTGCGGAGCGATCAAGGGCCAGGAG 3628749  
QY 869 CTCGCTTATCATCAGAACTCGACCCATCTGCGCTCTCCAGCTGCCATGGAGGCT 928  
Db 3628748 CCGGGGCTCCGTCACCGAGATCGACCCGATCAACCGCTGCGAGCGATGATGGAGGCT 3628689  
QY 929 ACCAGGTCGCCGATCGAGGAGTCTGTAAGGATTCGATATCTTCTGTTACATGACAG 988  
Db 3628688 TCGAGCTGGTCAACGCTGAGGAGGCGCATCGGGAACCGGACATGTCGTAACCGGACCG 3628629  
QY 989 GAATCTCGATATCTCTGTTGACATGATGGCCAGATGAAGATGAAGCTTATGTCG 1048  
Db 3628628 GCAACAAGACATCATCTGCTCGAGCACATTAAGCGGATGAAGGACACGCGATCTGG 3628569  
QY 1049 GTACATCGGCACTTCGATACGAATTTGATACATGGCTCATGAATATCCAGGCA 1108  
Db 3628568 GAATATCGGCGACTTTCGACACGAGATCGATGCGCGGCTGGAGCGCTCGGGGCA 3628509  
QY 1109 TCAAGCAATCCCAATCAAGCAGAAATACGACATGTGGAAATTCAGATGGCCAGCTTA 1168  
Db 3628508 CACGGGTCAACGCTCAAGGCTCAGGTCGACCTGTGGACCTTTGGGACACAGCGCGCTCGA 3628449  
QY 1169 TCCTCTCTTGTGAGGGCGCGCTTCTTAACCTTGGCTGGCTACAGGTCACCCATCTT 1228

Db 3628448 TCATCTGCTGTCGAGGGCGGCTGCTGATCTGCGCAATGCCCGGACCCCTGCT 3628389  
QY 1229 TCGTTATGTAATGTCATTCACAAACGACACTCGCTCAGCTCGACTTACGAAAAAGA 1288  
Db 3628388 TCGTGATGAGCAACAGCTTCGCTAACCAACGATGCCAGATCGCTGTGACCAAGA 3628329  
QY 1289 GAGGAATCTCCAGAGAGGTTTACACACTTCCGAGCATCTCGATGAGAGTGCCTC 1348  
Db 3628328 ACGAGGATGACACACAGAGGTGTACCGGCTGCCAGACCTCGAGAGAGTGGTCTC 3628269  
QY 1349 GCCTCCACCTTCGATCTCTCGATGTCACACTTACAAAGCTTACACAGAGAGGCTGACT 1408  
Db 3628268 GAATCCATGTGAGGCGCTTGGCGGTCACTGACCAAGTCAACAGGACGAGGCGAAT 3628209  
QY 1409 ACATCAACGTTCCAGTTCAGGCTCTTACAGTCTGATGCTTACCGTTATTA 1461  
Db 3628208 ACCTCGCGGTGACGCTGAGGCTCCCTACAGCGGACCTACCGCTACTGA 3628156

RESULT 12  
AAF71864  
ID AAF71864 standard; DNA; 1396 BP.  
XX AAF71864;  
AC  
XX  
DT 30-APR-2001 (first entry)  
XX  
DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:223.  
XX  
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;  
KW fine chemical production; microorganism; organic acid; nucleoside;  
KW nonproteogenic amino acid; purine base; pyrimidine base; nucleotide;  
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;  
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN WO200100843-A2.  
XX  
PD 04-JAN-2001.  
XX  
PF 23-JUN-2000; 2000WO-1B0923.  
XX  
PR 25-JUN-1999; 99US-0141031.  
PR 01-JUL-1999; 99DE-1030476.  
PR 02-JUL-1999; 99US-0142101.  
PR 08-JUL-1999; 99DE-1031415.  
PR 08-JUL-1999; 99DE-1031418.  
PR 08-JUL-1999; 99DE-1031419.  
PR 08-JUL-1999; 99DE-1031420.  
PR 08-JUL-1999; 99DE-1031424.  
PR 08-JUL-1999; 99DE-1031428.  
PR 08-JUL-1999; 99DE-1031434.  
PR 08-JUL-1999; 99DE-1031435.  
PR 08-JUL-1999; 99DE-1031443.  
PR 08-JUL-1999; 99DE-1031453.  
PR 08-JUL-1999; 99DE-1031457.  
PR 08-JUL-1999; 99DE-1031465.  
PR 08-JUL-1999; 99DE-1031478.  
PR 08-JUL-1999; 99DE-1031510.  
PR 08-JUL-1999; 99DE-1031541.  
PR 08-JUL-1999; 99DE-1031573.  
PR 08-JUL-1999; 99DE-1031592.  
PR 08-JUL-1999; 99DE-1031632.  
PR 08-JUL-1999; 99DE-1031634.  
PR 08-JUL-1999; 99DE-1031636.  
PR 08-JUL-1999; 99DE-1032125.  
PR 09-JUL-1999; 99DE-1032126.  
PR 09-JUL-1999; 99DE-1032130.  
PR 09-JUL-1999; 99DE-1032186.  
PR 09-JUL-1999; 99DE-1032206.  
PR 09-JUL-1999; 99DE-1032227.

```

PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032922.
PR 14-JUL-1999; 99DE-1032926.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1033004.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 12-AUG-1999; 99US-0148613.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040832.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041380.
PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.
XX
PA (BADI ) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI; 2001-137957/14.
DR P-PSDB; AAB79745.
XX
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
PT pathway proteins, useful for producing fine chemicals in
PT microorganisms, including organic acids, nonproteinogenic amino acids,
PT and purine and pyrimidine bases -
XX
XX Claim 3; Page 477-479; 173pp; English.
XX
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC pathway (NP) proteins given in ABE79634 to AAB80211. The C. glutamicum
CC NP nucleic acids are useful for the production of fine chemicals
CC in microorganisms, including organic acids, nonproteinogenic amino
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC compounds, vitamins, cofactors, polyketides and enzymes.
XX
XX Sequence 1396 BP; 309 A; 396 C; 384 G; 307 T; 0 other;
XX
Query Match 36.2%; Score 529.6; DB 22; Length 1396;
Best Local Similarity 64.3%; Pred. No. 7.3e-146;
Matches 828; Conservative 0; Mismatches 454; Indels 6; Gaps 2;
Y 32 TCGAGTACAGAAATCGCGACATCAACTCCATGTTCTCGGGCGTAAGAACTTACCCCTTG 91
D 114 TGGACTTCAGGTTGGCGATCTTCTACTACGAGCGAGGACGTCACCAATTCGCTTG 173
Y 92 CTGAGAGGAAATGCGAGTCTTATGTTCTTCGTGAGGTTATTCGGCTTCPAAGCAAT 151
D 174 CAGAGTATGAGATGCGAGGTCCTCATGCGATTGCGCAAGAAATCGCAGAGGAGCCTT 233
Y 152 TGAAGGTTGCAAACTCTGTTCCCTCCACATGACATCCACAGACGCGTCTCTACG 211
D 234 TGAAGGCGCCGAATGCTGTTCTATPCCACATGACGTTCCAGACCGCGTCTATTG 293
Y 212 AGACACTCACAGCTCTTGTTGATGATGTCAGATGGGCTTCCTGCAACATCTTCTACAC 271
D 294 AGACCTCACTGCTTTGGGCGCTGAGGTGCTGGGCTTCCTGCAACATTTCTCCACC 353

```

```

QY 272 AAGATACAGCCGCTGCTGCTATCTGTTCTGGGCCCAACAGGCACACAGAGACGACGCG 331
D 354 AGGATGAGGCTGCAGCGCTATCTGTTCTGGC---TCCGGCACCGTCGAAGACGAGTGTG 410
QY 332 GTATCCCAAGTCTTCGCTGGAAGGGCGAAACACTCCAGAAATATCTGGGAGAACATACC 391
D 411 GTGTTCCAGTATTCGGGTGGAGGGTGAGTCACTGGAGGAGTACTGTTGTTGATCAACC 470
QY 392 CGGCTCTCATATGCGCAATGCTCAAGGCCCAAGAGGTTGTGATGATGTTGTTGATG 451
D 471 AGATCTTCAGCTGGGGCGATG---AGCTGCCAAACATGATCTCGACGACGCGGTGAG 527
QY 452 CTACACTCTCTCATCTCCAAGGCTTCGAAATTCGAAACAGCGGCTGCTGTTCCAGAGCAA 511
D 528 CCACCATGGCTGTATTTCGGGTGCGGAATACGAGCAGGCTGTTCTGTTCCACAGCAG 587
QY 512 CAGAAGCTGCAACCTCGAATACCGCTGCGTTCCTTACACTCAAGCAGGTTCTCAACC 571
D 588 AGGCCAAACGATTCGATGAGTACATCGCATTTCTTGGCATGCTCGTGAGGTTCTTGTG 647
QY 572 AAGACAAGAACCACTGGGCACACAGTTCCTCGCGCATGAACGTTGTTCCGAGAGACAA 631
D 648 CAGAGCCTGCAAGTGGGGCAAGATCTCTGAGCCGTTAAGGGTGTCAACGAGAAACA 707
QY 632 CAAAGGTGTCACCGCTCTACAGCTCGAGAGGAGGCAAACTCTCTTCCAGACA 691
D 708 CCACCGGTGTGACACCGCTTACCACTTCCTGAAAGAGGCTGCTGCTTCCAGAGA 767
QY 692 TCAACGTCACAGAGCGCTGTTCAAAGTCCAAAGTTCGATAAATCTACGGTCGCGCACT 751
D 768 TGAACGTCACAGAGCGCTGTCAAGTTCGAGTTGATAACAGTAGGCACCCGCCACT 827
QY 752 CCCTTATCGATGATCAACCGTCTTCCGATGATCGCGGCAAGACAGTCTG 811
D 828 CCCTGATCGAGCGCATCAACCGCCACTGACATGCTCATGCGCGCAAGAGCTCTG 887
QY 812 TCATGGGTACGGGATGTCGGAGGCTGCGCTCAATCCCTCGTGCCAGAGGCTC 871
D 888 TCTGGGTACGGGATGTCGGAGGCTGCGCTGAGGCTTTGACGCGCAGGGGCTC 947
QY 872 GCGTTATCATCAGAACTCGACCAATCTGCGCTCTCCAGGCTGCCATGGAAGGCTAC 931
D 948 GCGTCAAGGTCACCGAAGCTGACCCAATCAACGCTCTTCAGGCTCTGATGGTGTACT 1007
QY 932 AGGTCGCGCATCGAGGAGTTCGATGAGGATGTCGATATCTTCGTTACATGACAGGAA 991
D 1008 CTGTGGTCAACGGTTGATGAGGCAATCGAGGACCGCCACATGTCATCCGCGACCGCA 1067
QY 992 ACTGCGATATCATCTCTGTGACATGATGCGCCAGATGAAGGATAAGGCTATTGTG 1051
D 1068 ACAAGGACATCAITTCCTTCGAGCAGATGCTCAAGATGAAGGATCAAGCTCTCTG 1127
QY 1052 AGATGCGCATTCGATAACGAATTCATACATGAGCTCATGAATACCGAGCATCA 1111
D 1128 ACATCGGTCACTTGATATGAGATGATATGATTCCTGTTCCACCGGACGAGCTCA 1187
QY 1112 AGCAGATCCCAATCAAGCCAGAAATACCAATGTGGAATTCACAGATGCGCAGCTATCC 1171
D 1188 CCGCCACCAAGTACAGTCCAGGTCGACGAGTTACCTTCTCCACCGGTCGCTCCATCA 1247
QY 1172 TCTTCTTGTGAGGGCGGCTTCTTAACCTTGGCTGCGCTACAGGTCACCGATCTTCG 1231
D 1248 TCGTCTGTCCGAAGTGGCTTGTGAACCTTGGCAACGCCACCGGACACCATCATTTG 1307
QY 1232 TTATGTCAATGTCAATTCACAAACAGACACTTCCTCAGCTCAGCTCTACGAAAGAGAG 1291
D 1308 TCAATGTCAACTCTTTCCCGGTTCAGACCAATTCGCGAGATCGAATCTGTTCCAAACAG 1367
QY 1292 GAAATCTCGAAGAGAGGTTTACACT 1319
D 1368 GACAGTACGAGACGAGGTTACCGTCT 1395

```

RESULT 13  
 AAS96123  
 ID AAS96123 standard; DNA; 1396 BP.  
 XX  
 AC AAS96123;  
 XX  
 DT 26-FEB-2002 (first entry)  
 XX  
 DE C. glutamicum gene #48 encoding metabolic pathway protein.  
 XX  
 KW Metabolic pathway protein; MP; lysine biosynthesis pathway;  
 KW methionine biosynthesis pathway; large-scale production of fine chemical;  
 KW Corynebacterium diphtheriae; diphtheria; ds.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN WO200166573-A2.  
 XX  
 PD 13-SEP-2001.  
 XX  
 PF 22-DEC-2000; 2000WO-IB02035.  
 XX  
 PR 09-MAR-2000; 2000US-187970P.  
 PR 23-JUN-2000; 2000US-060574D.  
 XX  
 PA (BADI ) BASF AG.  
 XX  
 PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberkauer G, Kim J;  
 PI Lee H, Hwang B;  
 XX  
 DR WPI; 2001-582269/65.  
 DR P-PSDB; AAU71913.  
 XX  
 PT Nucleic acids encoding metabolic pathway proteins from Corynebacterium  
 PT glutamicum, useful for producing methionine and lysine in  
 PT Corynebacterium and Brevibacterium -  
 XX  
 PS Disclosure; Page 281-283; 316pp; English.  
 XX  
 CC The present invention relates to the isolation of novel Corynebacterium  
 CC glutamicum genes encoding metabolic pathway (MP) proteins  
 CC (AAU71863-AAU71922). The metabolic pathway proteins of the invention  
 CC include enzymes involved in the lysine and methionine biosynthetic  
 CC pathways. The polynucleotide sequences of the invention can be used  
 CC for the large-scale production and/or modulation of expression of the  
 CC fine chemicals such as lysine and methionine. The sequences of the  
 CC invention may be used to identify C. glutamicum and related organisms  
 CC e.g. C. diphtheriae in a subject to detect diphtheria.  
 CC AA96073-AA96132 represent C. glutamicum genes encoding the novel  
 CC metabolic pathway proteins of the invention.  
 XX  
 SQ Sequence 1396 BP; 309 A; 396 C; 384 G; 307 T; 0 other;  
 Query Match 36.2%; Score 529.6; DB 23; Length 1396;  
 Best Local Similarity 64.3%; Pred. No. 7.3e-146;  
 Matches 828; Conservative 0; Mismatches 454; Indels 6; Gaps 2;  
 YQ 32 TCAGATACAGATTCGCGACATCACTCCATGTTCTCGCGCTAGGAACTACCTTG 91  
 DB 114 TGACCTTCAAGTTTCCGATCTTTCTAGCAGAGGAGGACGTCACAGATTCGCTTG 173  
 YQ 92 CTGAGAGGAAATGCCAGGCTTTATGGTTCTTCGTGAGCGTTATTCGGTTCTTAAGCCAT 151  
 DB 174 CAGAGTATGAGATGCCAGGTTCTCATGCGATTGCGCAAGGAATTCGACAGGAGCGCTT 233  
 YQ 152 TGAAGGTTGCAAGATCTCGGTTCCCTCCATGACAGTCCAGACAGCGCTTCATCG 211  
 DB 234 TGAAGGCGCCCGAAATTCCTGGTTCTATCCATGACGGTCCAGACGCGGCTTATTG 293  
 YQ 212 AGACACTCAGCTCTTGGTCTGATGTCAGATGGGCTTCTGCAACATCTTCTACAC 271  
 DB 294 AGACCTTCACTCTTGGGCGCTGAGGTTCTGTTGGGCTTCTGCAACATCTTCTACACC 353

YQ 272 AAGATACAGCGCTGCTGCTATCGTTTGGGCCCAACAGGCAACACAGAGGAGCGCG 331  
 DB 354 AGGATGAGGCTGACGCGCTATCGTTTGGGC---TCCGGCACCGTCCAGAGCCAGCTG 410  
 YQ 332 GTATCCCAAGTCTTCGCTGGAAGGCGAAACACTCCAGAACTACTGGGAGAACACATACC 391  
 DB 411 GTGTTCCAGTATTTCGGTGGAGGCTGAGTCACTGGAGGAGTACTGTGTGTCATCAACC 470  
 YQ 392 GGGCTCTCATGCGCCAGATGTCAGGCCCCACAGAGGTTGTGATGATGATGTTGGTATG 451  
 DB 471 AGATCTTCAGCTGGGCGATG---AGTGCACAAACATGATCTCGACAGCGGCGTGAAG 527  
 YQ 452 CTACACTCTCTCATCTCCAGGGCTTCGAATTCGAACACAGCGGCTGCTGTTCCAGAGCAA 511  
 DB 528 CCACCATGGCTGTTATTTCGGGTGCGGAATACGAGCAGGCTGGTCTGGTTCACAGCAG 587  
 YQ 512 CAGAAGCTGACAACTCTGAAATACCGCTGGCTTCTTCTTACACTAAGCAGTCTTCAACC 571  
 DB 588 AGGCCAAACGATTCCGATGAGTACATCGCAATTCTTGGGCATGCTGCGTGAGGTTCTTGTG 647  
 YQ 572 AAGCAAAAGAACCACTGGCAACAGTTGTCGCGCATGAACGGTGTGTTCCGAAGAGACAA 631  
 DB 648 CAGAGCTGCAAGTGGCGCAGATCGCTGAGSCCTTTAAGGTTGTCAACGAGAAACCA 707  
 YQ 632 CAACAGGTGTCACCGCTCTTACAGCTCGAAGAGGAGGCAAACTCTTCTCCAGGCA 691  
 DB 708 CCACCGGTGTGCAACCGCTTACCACTTCGCTGAAGAGGCGTGTGCTTTCACAGCA 767  
 YQ 692 TCAACGTCACAGCGCTGTACAAAGTCCAAAGTTCAGATTAAGTCTAGCGCTGCGGCACT 751  
 DB 768 TGAACGTCACAGCGCTGTACAAAGTCCAAAGTTCAGATTAAGTCTAGCGCTGCGGCACT 827  
 YQ 752 CCTTATTCGATGATCAACCGTGTTCGATGTCTATGATCGGCGCAAGCAGCTCTCG 811  
 DB 828 CCTGATCGAGCGCATCAACCGCGCACTGACATGCTCATGGCGGCAAGAACGCTGTTG 887  
 YQ 812 TCATGGGTTACGGGATGTGCGCAAGGCTGCGCTCAATCCCTCGGTGGCAAGGGGCTC 871  
 DB 888 TCTCGGTTACGGGATGTGCGCAAGGCTGCGCTTTCGAGCGGCAAGGGGCTC 947  
 YQ 872 GGGTTATCATCAGAACTCCACCAATCTGCGCTCTCCAGGCTGCCATGGAAGGCTACC 931  
 DB 948 GGGTCAAGGTCACCGAAGCTGACCCCAATCAACGCTCTTCAGGCTCTGATGATGCTACT 1007  
 YQ 932 AGGTCGCGGCTCTGAGGAGTGTCAAGATGTGATATCTTGTACATGACAGGAA 991  
 DB 1008 CTGTGGTCAACCGTTGATGAGGCAATCGAGGACCGGACATCGTATCAGCGGCAAGGCA 1067  
 YQ 992 ACTCGATATCATCTCTGTGACATGATGCCAGATGAAGGATAAGGCTATTGTGGTA 1051  
 DB 1068 ACAAGGACATCATCTCTTCGAGCAGATGCTCAAGATGAAGATCACTCTCTCTCTGCGGA 1127  
 YQ 1052 ACATCGGCGCTTCGATACGAATGATACAGATGGCTCATGAAATACCCAGGATCA 1111  
 DB 1128 ACATCGGCTCATCTTGATATGAGATCGATATGCAATTCCTGTTGACCGGCAAGGCACTCA 1187  
 YQ 1112 AGCAGATCCCAATCAGGCAAGATACGACATGCGGAATTCAGATGGCGCAGCTATCC 1171  
 DB 1188 CCGGACCAAGATCAAGCCACAGGTCGAGGATTCACCTTCTCCACCGGTCGCTCCATCA 1247  
 YQ 1172 TCCTTCTTGTGAGGCGGCTCTTAACTTGGCTGCGCTACAGGTCACCCATCTTTCG 1231  
 DB 1248 TCGTCTGTCCGAAGTGTGCTTTGAACCTTGGCAACGCCACCGGACACCCATCATTTG 1307  
 YQ 1232 TTATGTCAATGTCATTCACAAACAGACACTCGCTCAGCTGACCTCTACGAAAGAGAG 1291  
 DB 1308 TCATGTCCAACTCTTTGCGCGATCAGACCAATTCGAGATCGAACTGTTCGAAAGAGAG 1367  
 YQ 1292 GAAATCTCGAGAGAGGTTTACACT 1319  
 DB 1368 GACAGTACGAGAGGTTTACCGTCT 1395

RESULT 14  
AAC44037  
ID AAC44037 standard; DNA; 1865 BP.  
XX AAC44037;  
AC AAC44037;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Zea mays DNA fragment SEQ ID NO: 41385.  
XX  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway; metabolic;  
KW pathway; promoter; termination sequence; corn; ss.  
XX  
OS Zea mays subsp. mays.  
XX  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0132865.  
PR 14-MAY-1999; 99US-0134216.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135355.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137500.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.



PR	20-AUG-1999;	99US-0149723.	99US-0149929.	214	ACACTCAAGCTCTTGGTGTGATGTCAGATGGCTTCCTGCAACATCTTCTTACACA	273
PR	23-AUG-1999;	99US-0149902.	99US-0149902.	293	ACCTTACCGCGCTCGGCGGAGGTGCGTGTCTCTGCAACATCTTCTCCAGCAG	352
PR	25-AUG-1999;	99US-0150566.	99US-0150566.	274	GATACAGCCGCTGCTCTATCTGTTGTGCGCCCAACAGGCACACCAGAGAAAGCCAGCGGT	333
PR	26-AUG-1999;	99US-0150884.	99US-0151065.	353	GACCACGCCGCCGCCATCCGCGCGACTCGGC-----	387
PR	27-AUG-1999;	99US-0151066.	99US-0151080.	334	ATCCAGTCTTCCTGGAAGSGGAAACACTCCAGATACTGGGAGAACACATCCCGC	393
PR	31-AUG-1999;	99US-0151303.	99US-0151436.	388	--CGCGTGTTCCTGGAAGSGGAGACCTTTAGGAGTACTGGTGTGACCGAGC	445
PR	01-SEP-1999;	99US-0152363.	99US-0152363.	394	GCTCTCAATGCGCCAGATGCTCAAGGCCACAGAGTTGTCGATGATGATGATGATGAT	453
PR	10-SEP-1999;	99US-0153070.	99US-0153758.	446	TGCTTGAATGGGCGAGCGCGGCGGCCGACCTCATGTCGACGCGGCGGCGCGC	505
PR	13-SEP-1999;	99US-0153758.	99US-0154018.	454	ACACTCTCTCTCTCAAGGGCTTC-----GAATTCGAAACAGCGCGTGTGT	501
PR	16-SEP-1999;	99US-0154039.	99US-0154779.	506	ACGCTGTCTCTCAAGGGTGTCAAGCGCGAGGAGGTACGAGAAAGCCGCAAGATC	565
PR	20-SEP-1999;	99US-0154779.	99US-0155139.	502	CCAGAGCCAAACAGAGCTGACAACTCGAATACCGTGGTTCCTTGTCTACACTCAAGCAG	561
PR	23-SEP-1999;	99US-0155486.	99US-0155486.	566	CCGACCGGAGTCCACGACAAAGCTGATGATGATGATGATGATGATGATGATGATGAT	625
PR	28-SEP-1999;	99US-0156458.	99US-0156596.	562	GTCTTCAACCAAGACAAAGAACCTTGGCACACAGTTGTGCGCGCATGAACGGTGTTC	621
PR	04-OCT-1999;	99US-0157117.	99US-0157753.	626	GGGCTCAAGGCTGACCCCAAGAGTACCGCAAGTGAAGGAGGCTTGTGCGGCTCT	685
PR	06-OCT-1999;	99US-0157865.	99US-0159293.	622	GAGAGCAACACAGGTTGTCACCGGCTCTACAGCTCGAGAGGAGGCAACTCTC	681
PR	08-OCT-1999;	99US-0158029.	99US-0158232.	686	GAGAGACCCACCGGTTGTCAGAGGCTCTACAGATGAGGAGACCGGCGCCCTCTC	745
PR	12-OCT-1999;	99US-0158369.	99US-0159293.	682	TTCCAGCCATCAAGCTCAACGAGCTGTTTACAAAGTCCAAAGTTCGATTAACATCTACGC	741
PR	13-OCT-1999;	99US-0159294.	99US-0159331.	746	TTCCCTGCCATTAAGTCAACGATTCGTCACCAAGCAAGTTTGACAACTGTATGT	805
PR	13-OCT-1999;	99US-0159295.	99US-0159332.	742	TGCGGCACTCCCTTATCGATGATGATCAACGCTGCTCCGATGTCATGATGCGGCGCAG	801
PR	14-OCT-1999;	99US-0159330.	99US-0159337.	806	TGCGGCACTCACTCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	865
PR	14-OCT-1999;	99US-0159330.	99US-0159338.	802	ACAGCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	861
PR	18-OCT-1999;	99US-0159584.	99US-0160741.	866	GTGCGGCTGCTCGGATACGGTGTGTCGCAAGGTTGTCGCGTGCCTCAAGCAG	925
PR	21-OCT-1999;	99US-0160741.	99US-0160768.	862	CAAGCGCTCGGTTATCATACAGAACTCGACCCAACTGCTCTCCAGGCTGCTCAGT	921
PR	21-OCT-1999;	99US-0160815.	99US-0160980.	926	GCTGTCGCGCTGCTCATTTGTGACCGAGATCGACCCCATCTGCGGCTCCAGGCTCTGATG	985
PR	22-OCT-1999;	99US-0160981.	99US-0160989.	922	GAAGGCTACAGGTCGCGCATCGAGAAAGTCTCAAGGATGTCGATATCTTCGTTACA	981
PR	25-OCT-1999;	99US-0161404.	99US-0161405.	986	GAGGCTTTCAGGTCCTTCCCTTGGAGGAGTGTCTCGAAGTGCATCTTCGTGACC	1045
PR	25-OCT-1999;	99US-0161406.	99US-0161359.	982	TGCAAGGAACTGCGATATCATCTCTGTTGACATGATGCGCCAGATGAAGGATAAGGT	1041
PR	26-OCT-1999;	99US-0161360.	99US-0161361.	1046	ACCACTGGCAACAGGATATCATCTGTTGACCACTGAGGAAGATGAAGAACATATGC	1105
PR	28-OCT-1999;	99US-0161920.	99US-0161992.	1042	ATTGTCGCTAACATCGGCGCTTCGATAAGAAATGATACAGATGCGGCTCATGAATAC	1101
PR	28-OCT-1999;	99US-0161992.	99US-0161993.	1106	ATTGTCGCAACATGCGGCTTTCGACATGAATGATGATGATGATGATGATGATGATGATGAT	1165
PR	29-OCT-1999;	99US-0162142.	99US-0162142.	1102	CCAGGCTAACGACATCCCAATCAAGCCAGAAATCGACATGTGGGAATTCGCCAGATGCC	1161
PR				1166	CCTGGGCTCAAGGCTATCACCATCAAGCCAGGCTGCGGCTGGGTTTCCCGAGACC	1225
PR				1162	CAGGCT---ATCCTCTCTTCTGCTGAGGCGGCTTCTTAACTTGTGCTGCGTCAAGGT	1218
PR				1226	AACACTGGCATCATTTGCTCTGCTGAGGCTGCGGCTGATGAACCTTGGGTGTGCTACTGCG	1285
PR				1219	CACCACTTTTCTGTTATGTCATGTCATTCACAAACAGACACTCGCTCGCTCAGCTC	1278
PR				1286	CATCTAGCTTTGTCATGTCCTGCTCATTCATACAGGTCATTTGCCCAACTTGAAGT	1345

Query Match	34.8%;	Score	508.8;	DB	21;	Length	1865;
Best Local Similarity	62.0%;	Pred. No.	1.2e-139;				
Matches	898;	Conservative	0;	Mismatches	502;	Indels	48;
Gaps	4;						
QY	34	GAGTACAGAATCCGACATCAACCTCCATGTTCTCGGCGGTAAAGAACTTACCTTGTCT	93				
Db	113	GAGTACAGGTCAAGATCTCTCGAGGCGGACTTCGCGCGCTCGAGATTGAGCTGGCC	172				
QY	94	GAGAGGAATTCGACGCTTATGTTCTTCGTGAGCGTTATTCGCTTCTTAAGCAATGG	153				
Db	173	GAGTCAAGATCCCGGCTCATGCGTCCCGCGCGAGTTCCGCGCTTCAGAGCCTTC	232				
QY	154	AAGGTTGTCAGATCTCTGTTTCCCTCCCATGACAGTCCAGACAGCGCTCTCATCGAG	213				
Db	233	CGCGCGCTAGAGATCTCGGGGTCTCTCCATGACCATCCAGACCGCGCTCTCATCGAG	292				

```

QY 1279 TACGAAAGA-----GAGGAATCTCGAGAAAGGTTTACACTTCCGAGCATCTC 1332
Db 1346 TCGAAGGAGAGAGCTCTGGCAAGTATGAGAAAGGTTGTTGCTCCCAAGCACTT 1405
QY 1333 GATGAGAAGTCTGGCTCCACCTCGGATCTCTCGATGTCACCTTACAAGCTTACA 1392
Db 1406 GATGAGAAGTCTGGCTCTCCACTTGGCAAGCTTGGTCCCAAGCTGACCAAGCTCAC 1465
QY 1393 CAGAGCAGGCTGACTACATCAACGTTCCAGTTGAGGTCCTTACAAAGTCTGATCTTAC 1452
Db 1466 AATCTCAGGCGCACTATCATCGGTGCGGATCGAGGTCCTTACAAAGCTGCCACTAC 1525
QY 1453 CGTTATTA 1460
Db 1526 CGGTACTA 1533

RESULT 15
AAC46315
ID AAC46315 standard; DNA; 1727 BP.
XX AC AAC46315;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 49689.
XX KW Hybridisation assay; genetic mapping; gene expression control;
XX KW protein identification; signal transduction pathway;
XX KW metabolic pathway; promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 18-JUL-1999; 99US-0144325.
PR 18-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.

```



Qy	1042	ATTGTCGGTAACATCGGCCCTTCGATACGAAATGTACAGATGCCTCATGAATAC	1101
Db	1099	ATTGTCGCAACATGGTCACTTTGACAAATGAGATTGACATGTGGACTTGAGACTTAC	1158
Qy	1102	CCAGGCATCAAGCACATCCCAATCAAGCCAGAAATACGACATGTGGAAATCCAGATGCC	1161
Db	1159	CTTGGTGAAGCGTATCACCATCAAGCCACAGACTGACAGTGGGTGTGCCAGAGACC	1218
Qy	1162	CACGCT---ATCCTCCTTCTGTGAGGGCCGCTTCTTAACCTTGGCTGGGCTACAGGT	1218
Db	1219	AAGGCTGGAATCAATGCTTGGCTGAGGGTCGTCTGATGAACCTGGGTGTGCCACTGGT	1278
Qy	1219	CACCCATCTTTCGTTATGTCAATGTCATTCACAAACACAGACACTCCCTCAGCTCGACCTC	1278
Db	1279	CACCCAAAGTTGGTGATGCTGTCTTTACCCACCAGGTGATGCCCCAGCTCGAGCTC	1338
Qy	1279	T-----ACGAAAAGAGAGGAATCTCGAAGAGAGGTTTACACACTTCGAGGATCTC	1332
Db	1339	TGGAACGAGAAGCAAGCGGAAGTAGCGAGAGAGGTGTACGTTCTTCCAGCATTTG	1398
Qy	1333	GATGAAGAGTCGCTCGCCTCCACCTCGGATCTCTCGATGTCACCTTACAAAGCTTACA	1392
Db	1399	GATGAAGAGTTGCATTACTTCACTTGGGCAAGCTTGGAGCCAGGCTTACAAAGCTGTCA	1458
Qy	1393	CAGAAGCAGGCTGACTACATCAACGTTCCAGTTGAGGGTCCTTACAGTCTGATGCTTAC	1452
Db	1459	AAGGACCAATCTGACTAGTCAGCATTCCAATTGAGGACCATTACAAGCCTCTCTACTAC	1518
Qy	1453	CGTTATTAA	1461
Db	1519	AGTACTGA	1527

Search completed: October 17, 2003, 20:27:06  
Job time : 471 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2003, 20:19:45 ; Search time 104 seconds

(without alignments)  
6200.584 Million cell updates/sec

Title: US-09-759-990A-1

Perfect score: 1461

Sequence: 1 atggcttgcaatcacctac.....ctgatcttaccgttattaa 1461

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PTUS.COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	535.4	36.6	4403765	3	US-09-103-840A-2
C 2	535.4	36.6	4411529	3	US-09-103-840A-1
C 3	490	33.5	1812	2	US-08-669-536-1
C 4	489	33.5	1767	3	US-08-930-894-1
C 5	223.2	15.3	2211	3	US-09-318-448-26
C 6	223.2	15.3	2211	4	US-09-347-878-2
C 7	179.4	12.3	1485	4	US-09-252-991A-6815
C 8	179.4	12.3	1509	4	US-09-252-991A-6930
C 9	167	11.4	1392	4	US-09-328-352-474
C 10	153.8	10.5	288	4	US-09-702-705-1179
C 11	153.8	10.5	288	4	US-09-736-457-1179
C 12	122.4	8.4	289	1	US-08-204-740-8
C 13	122.4	8.4	289	3	US-09-081-167A-8
C 14	122.4	8.4	289	3	US-09-081-395-8
C 15	122.4	8.4	289	3	US-09-416-833-8
C 16	122.4	8.4	289	5	PCT-US95-02521-8
C 17	114.6	7.8	1664976	4	US-08-916-421B-1
C 18	111.8	7.7	648	4	US-09-252-991A-6726
C 19	105	7.2	285	1	US-08-204-740-6
C 20	105	7.2	285	3	US-09-081-167A-6
C 21	105	7.2	285	3	US-09-081-395-6
C 22	105	7.2	285	3	US-09-416-833-6
C 23	105	7.2	285	5	PCT-US95-02521-6
C 24	105	7.2	966	4	US-09-252-991A-6895
C 25	96.6	6.6	584	3	US-08-998-416-249
C 26	80.8	5.5	299	4	US-09-313-294A-4325
C 27	56.8	3.9	636	3	US-09-328-111-850

C 28	52.8	3.6	309	4	US-09-252-991A-6725	Sequence 6725, Ap
C 29	42.8	2.9	3379	4	US-09-489-847-27	Sequence 27, Appl
C 30	42.2	2.9	537	4	US-09-252-991A-2497	Sequence 2497, Ap
C 31	42.2	2.9	2088	4	US-09-252-991A-2576	Sequence 2576, Ap
C 32	41.8	2.9	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 33	41.8	2.9	4411529	3	US-09-103-840A-1	Sequence 1, Appli
C 34	41	2.8	390	3	US-09-197-649-7	Sequence 7, Appli
C 35	41	2.8	1125	4	US-09-252-991A-11272	Sequence 11272, A
C 36	41	2.8	1923	4	US-09-252-991A-10971	Sequence 10971, A
C 37	40.2	2.8	390	3	US-09-197-649-7	Sequence 7, Appli
C 38	40.2	2.8	3573	4	US-09-252-991A-4903	Sequence 4903, Ap
C 39	40	2.7	2226	2	US-08-896-005-2	Sequence 2, Appli
C 40	40	2.7	2226	4	US-09-347-878-3	Sequence 3, Appli
C 41	39.8	2.7	1251	4	US-09-252-991A-11883	Sequence 11883, A
C 42	39.8	2.7	1851	4	US-09-252-991A-11566	Sequence 11566, A
C 43	39.6	2.7	1551	4	US-09-252-991A-5309	Sequence 5309, Ap
C 44	39.6	2.7	1563	4	US-09-252-991A-15934	Sequence 15934, A
C 45	39.6	2.7	1842	4	US-09-252-991A-4840	Sequence 4840, Ap

#### ALIGNMENTS

##### RESULT 1

US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 2

; LENGTH: 4403765

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match.	36.6%;	Score 535.4;	DB 3;	Length 4403765;
Best Local Similarity	61.3%;	Pred. No. 2.3e-152;		
Matches 879;	Conservative	0;	Mismatches 551;	Indels 3; Gaps 1;
QY	32	TCGAGTACAGAAATTCGCCACATCAACCTCCATGTTCTCGGCGGTAAAGGAACCTTACCCCTTG	91	
Db	3624437	TCGACTTTAAGATCGCGACCTGTCACTAGCGGATTCGGCGCAAGAACTCCGGATCG	3624378	
QY	92	CTGAGAAGAAATGCCAGGTCTTATGGTTCTTCGTGAGCGTTATTCGGTCTCTAAGCCAT	151	
Db	3624377	CCGAGCAGGAGATCGCGGGCTGATGTCGCTCGCGCGAGTATGCCGAGGTGCAACCCC	3624318	
QY	152	TGAAGGGTGTGAGAATCTCTGGTTCCCTCCACATGACAGTCCAGACAGCGCTCTCATCG	211	
Db	3624317	TGAAGGGGCGCGGATCTCGGGTTCGTCGACATGACGGTGCAGACCGCGGTGTTCGATCG	3624258	
QY	212	AGACACTCAGAGCTTTGGTCTGATGTCAGATGGGCTTCCTGCAACATCTTCTCTACAC	271	
Db	3624257	AAACCTTCACCGCTGGGCGCGAAGTCCGCTGGGCTCGTGAACATCTTCTCTCACCC	3624198	
QY	272	AAGATACAGCGCTGCTGCTATCGTTTCGGGCCCAACAGGCAACACAGAGAAGCCAGCG	331	
Db	3624197	AGGATACGCGCGCGCGCTGTGGTTCGGCGCGACGACCGCCCGACGAGCAAGG	3624138	

```
QY 332 GTATCCAGCTCTGCGCTGAGGGCGGAAACACTCCAGAAATACCTGGGAGACACATACC 391
Db 3624137 GTGTCCTCCGGTGTTCGCTGAGAGGCGAGACGCTCGAAGAGTACTGCTGGCGCGCCGAGC 3624078

QY 392 GCGCTCTCACATGCGCCAGA---TGGTCAAGGCCACACAGAGGTGTTCGATGATGCTGGTG 448
Db 3624077 AGATGCTCACTGCTGCGGACCCCGACAGCCGCGCAACATGATCTCGATGACGCGGTG 3624018

QY 449 ATGTACTACTCTCACTCACTCAAGGGGTTTCGAATTCGAAACAGCGCGGTGCTGTTCAGAGC 508
Db 3624017 AGCCGACCATTTGGTGTCTGCGCGGCATGCGATGAGAAAGCGCGGCTGTGTCGCGCCG 3623958

QY 509 CAACAGAGCTGACAACTCGAATACCGCTCGGTTCTTGCTACACTCAAGCAGGTCTTCA 568
Db 3623957 CCGAGGAGGACGACCCCGCGAGTGAAGGCTCTTCTGAACTGCTACGACCGGCTTCG 3623898

QY 569 ACCAAGACAAGAACCACTGCGACACAGATTTGCTGCGGCATGAACGGTGTTCGGAAGAGA 628
Db 3623897 AGACCGACAAGGACAAGTGAACAAGATAGCGAGTCGGTCAAGGGCGTCAACCGAGAGA 3623838

QY 629 CAAACAAGGTGTCACCGCTCTACAGCTCGAAGAGGAGGCAAACTCTCTTTCCAG 688
Db 3623837 CCACACCGCGGTGCTGCGGCTCTACCAATTGCGCGCGCGCGGATCTGCGCTTCCCG 3623778

QY 689 CCATCAAGCTCAACGAGCTGTGTACAAAGTCCAAAGTTCGATACATCTAGCGCTGCCGCC 748
Db 3623777 CGATCAAGCTCAACGACTCGGTGACCAAGTCCAAATTTCGAAACAAAGTACGCGACTCGC 3623718

QY 749 ACTCCCTTATCGATGGTATCAACCGTGTCTCCGATGTCATGATCGCGCGCAAGACAGTCC 808
Db 3623717 ACTCCCTGATCGACGGCATCAACCGCGCACCGAGCGCTGATCGCGGTGAAGAGTCC 3623658

QY 809 TGCATAGGGTTACGGGATGTCGCAAGGGCTCGCTCAATCCCTCGTGGCCGAGCGG 868
Db 3623657 TCATTCGGGTACGGCGACGTCGCTGAAGGGCTGTGCGGAGGCGATGAAGGCGCGAGGAG 3623598

QY 869 CTCGCGTTATCATACAGAACTCGACCCAACTCGCGCTCTCAGAGGTGCGATGAAGGCT 928
Db 3623597 CGCGGGTCTCGGTACCGAGATCGACCGATCAACGGCTGCGAGGCCATGATGAGGGCT 3623538

QY 929 ACCAGGTGCGCGCATCGAGGAAGTCGTCAAGGATGTCGATATCTTCTGTTACATGCAAG 988
Db 3623537 TCGACGTGGTCACCGTCGAGGAGGCCATCGGGGACCGCGACATCGTCGTAAACCGGACG 3623478

QY 989 GAACTCGGATATCATCTCTGTGATGATGATGGCCGAGATGAGAGTAAGGCTATTGTG 1048
Db 3623477 GCAACAAAGACATCATCTGTCGAGCAATTAAGGGCGATGAAGGACCAACGCGATCTCG 3623418

QY 1049 GTAACTCGGCACTTCGATACGAAATTTGATACAGATGGCTTCATGAAATACCGAGCA 1108
Db 3623417 GAAATATCGGCCACTTCGACACGAGATCGACATGGCGGGGCTGGAGCGCTCCGGGGCG 3623358

QY 1109 TCAAGCACATCCCAATCAAGCCAGAAATACGACATGTGGGAATTCGCCAGATGGCCACGCTA 1168
Db 3623357 CACGGGTCAACGCTCAAGCTCAGTTCGACCTGTGACCTTTTGGCGACACGCGCGCTCGA 3623298

QY 1169 TCCTCTCTTCTGTGAGGGCGGCTCTTAACCTTGCCTGCGCTACAGGTACCGATCTT 1228
Db 3623297 TCATCTGCTGTGTCGAGGGGGGCGGCTGCTGAACCTGGGCAATGCGACCGGGCACCGCTCGT 3623238

QY 1229 TCGTTATGTCATGTCAATTCACAAACCAAGACACTCGCTCAGCTCGACCTCTACGAAAGA 1288
Db 3623237 TGTGATGAGCAACAGCTTCGCTTACCAAGAGATCGCCAGATCGAGCTGTGGACCAAGA 3623178

QY 1289 GAGGAATCTTCGAGAAAGAGTTTACACTTCCGAGAGCATCTTCGATGAAGAGTTCGCTC 1348
Db 3623177 ACGAGGATGACGACACGAGGTGTACCGGCTGCCCAAGACCTTCGACGAGAGGTGGCTC 3623118

QY 1349 GCTTCACCTCGGATCTCTGATGTCACCTTACAAAGCTTACAGAAAGAGGCTGACT 1408
Db 3623117 GAATCCATGTGAGGGCCCTTGGGGGTTCACCTGACCAAGCTGACCAAGAGGAGGCGCAAT 3623058

QY 1409 ACATCAAGTTCCAGTTGAGGGTCTTACAGTCTGATGCTTACCGTTATTAA 1461
```

```
Db 3623057 ACCTCGCGTCGACGTCGAGGGTCCCTCAAGCCGAGCACACTACCGCTACTGA 3623005

RESULT 2
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 36.6%; Score 535.4; DB 3; Length 4411529;
Best Local Similarity 61.3%; Pred. No. 2.3e-152;
Matches 879; Conservative 0; Mismatches 551; Indels 3; Gaps 1;

QY 32 TCGAGTACAGAAATGCGGACATCAACCTCCATGTTCTCGGCGGTAGGAACCTTACCGTTG 91
Db 3629588 TCGACTTTAAAGATCGCGGACCTGTCACTAGCGGATTTGCGCGCAAGAACTTCCGGATCG 3629529

QY 92 CTGAGAAGGAATGCCAGGTCTTATGTTCTTCTGAGCGTTATTCCGCTTCTAAGCCAT 151
Db 3629528 CCGAGCAGAGATGCCCGGCTGATGTCGTGCGCGCGATGATGCCGAGGTGCACACCC 3629469

QY 152 TGAAGGGTGTGAGAAATCTCTGTTCCCTCCACATGACAGTCCAGACAGCGTCTCATCG 211
Db 3629468 TGAAGGGGCGCGGATCTCGGTTCTGCTGCACATGACGTTGAGAGTCTGTTGATCG 3629409

QY 212 AGACACTCACAGCTCTTGTGCTGATGTCAGATGGGCTTCTGCAACATCTTCTACAC 271
Db 3629408 AAACCCCTCACCGGCTGGGCGCGGAGTCCGCTGGGCTCGTGCAACATCTTCTCCACC 3629349

QY 272 AAGATACAGCGCTGCTGCTATGTTGTGCGGCCCAACAGGCAACAGAGAAGCCAGCG 331
Db 3629348 AGGATCAGCGCGCGCGCGCGCTGCTGTCGCGCGCGACGCGCACCCCGAGAGCCAGG 3629289

QY 332 GTATCCGAGTCTTCGCTCGAAGGGCGAAACACTCCCAAGATACTGGGAGAACATACC 391
Db 3629288 GTGTCCCGTGTTCGCTGGAAGGGCGAGACGCTCGAAGAGTACTGTTGGCGCGCGAGC 3629229

QY 392 GGGCTCTCATGCGCAGA---TGGTCAAGGCCACAGCAGGTGTTCGATGATGCTGGTG 448
Db 3629228 AGATGCTCACTGCGCGGACCCCGGACAGCGGCAACATGATCTCTGATGACGCGGTG 3629169

QY 449 ATGCTACACTCTCATCTCCAAGGGCTTCGAATTCGAAACAGCGCGTCTGTTCCAGAGC 508
Db 3629168 AGCCCAACATGTTGGTGTGTCGCGGCATGAGTATGAGAGCGCGGTGGTGGCGCCG 3629109

QY 509 CAACAGAGCTGACAACTCTGAATACCGCTGCGTTCTTCTACACTCAAGCAGGTCTTCA 568
Db 3629108 CCGAGGAGGACGACCCCGCGAGTGAAGGTCTTCTCTGAACCTGCTACGACCGCTTGG 3629049

QY 569 ACCAAGACAAGAACCACTGCGACACAGTTCGTGCGGCGATGAACGCTGTTTCCGAAAGA 628
Db 3629048 AGACCGACAAGGACAAGTGAACCAAGATAGCGGATCGGTCAAGGGCGTCAACCGAGGA 3628989

QY 629 CAACAAAGGTGTCACCGCTCTACAGCTCGAAGAGGAGGCGCAAACTCTCTTCCAG 688
```

```
Db 3628988 CCACACCGCGTCTCGGCTCTACCAATTCGCGCGCGGGGATCTGGCCTTCCGG 3628929
Qy 689 CCATCAACGTCACACGCGTGTGTACAAAGTCOAAGTTCGATACATCTACGCGTCCGCGC 748
Db 3628928 CGATCAACGTCACACGCGTGTGTACAAAGTCOAAGTTCGATACATCTACGCGTCCGCGC 3628869
Qy 749 ACTCCCTTATCGATGATCAACCGTGTTCGATGTCATGATCGCGGCAAGACGCTC 808
Db 3628868 ACTCCCTTATCGATGATCAACCGGCGACCGCGCTGATCGCGGCTAAGAGTCC 3628809
Qy 809 TCGTCATGGGTATACGCGATGTCGCAAGGCTGGCTCAATCCCTCGTGGCCAAAGCG 868
Db 3628808 TCATCTGGGCTACGCGAGCTCGGTAGGCTGTGGAGGCGATCAAGGCGCGAGGAG 3628749
Qy 869 CTCGGGTATCATCAAGAACTCGACCCAACTCGGCTTCAGGTCGCAATGGAAGCT 928
Db 3628748 CGCGGCTCTCGCTACCGAGATCGACCCGATCAACGCGCTCGAGGCGATGATGAGGCT 3628689
Qy 929 ACCAGTCCGCGCATCGAGGAGTCTCAAGGATGTCAGATGTCGATATCTTCGTATCATGCAAG 988
Db 3628688 TCAGTGTGTCAGGAGGCGATCGGAGCGCGCATCTGCTGTAACCGGCGCG 3628629
Qy 989 GAACTGGCATATCATCTGTTGATCATGATGGCCAGATCAAGGATTAAGGCTATTGTG 1048
Db 3628628 GCAACAAGACATCATCTCTCGACATTAAGCGATGAAGACCGGATCTTGG 3628569
Qy 1049 GTAACTGGGCTCTCGATCAAGAACTGATCAAGATGCGCTCATGAATACCGAGCA 1108
Db 3628568 GAAATATCGGCACTTCGACACAGAGATCGATGGCGGCTGAGCGCTCCGGGCGA 3628509
Qy 1109 TCAAGCATATCCCAATCAAGCCAGATACGATCTGGGAATTCGCCAGATGSCACGCTA 1168
Db 3628508 CAGCGGTCAAGTCAAGCTTCAAGTCTGATGATGATGCGGCGGCTTGGCGACACGGCGGCTCGA 3628449
Qy 1169 TCTCTCTTCTGTCGAGGCGCGCTTCTTAACCTTGGCTGCGCTACAGGTCAACCATCTT 1228
Db 3628448 TCATCTGCTCTCGAGGCGCGCTCTGTAACCTGGCAATGCCACCGGCGACCCCTCGT 3628389
Qy 1229 TGTATGTCATCTTATCAAAACGACACTGCTGCTAGCTCGACCTCTAGAAAGA 1288
Db 3628388 TGTGATGAGCAACAGCTTCTGCTAACACAGCATGCGCCAGATGCGCTGTGACCAAGA 3628329
Qy 1289 GAGGAATCTCGAAGAAAGTTTACACACTTCCGAGCATCTCGATGAAGAGTCTGCTC 1348
Db 3628328 ACAGCGATGACGACAGAGGTGTCGCGCTGCCAAGACCTTCGACGAGAGGTGCTC 3628269
Qy 1349 GCTCCACCTCGATCTCTCGATGTCACCTTACAAAGCTTACAGAAAGCGAGGTGACT 1408
Db 3628268 GAATCATGTCGAGGCGCTTGGGCTACCTGACCAAGCTGACCAAGGAGCGGCGAAT 3628209
Qy 1409 ACATCAAGTTCAGGTGAGGCTCTTACAGTCTGATGCTTACCGTTATTA 1461
Db 3628208 ACTCGGCTGACGTCGAGGCTTCCCTACAAGCGGACCACTACCGCTACTGA 3628156

RESULT 3
US-08-669-536-1
; Sequence 1, Application US/08669536
; Patent No. 5910444
; GENERAL INFORMATION:
; APPLICANT: MASUTA, CHIKARA
; APPLICANT: UHARA, KYOKO
; APPLICANT: TANAKA, HIDEO
; APPLICANT: KUWATA, SHIGERU
; TITLE OF INVENTION: ORGANISMS IN WHICH THE EXPRESSION OF
; NUMBER OF SEQUENCES: 4
; S-ADENOSYLMOCYSTEINE HYDROLASE GENE IS INHIBITED
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
```

```
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent-In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,536
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1254-128
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1812 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; US-08-669-536-1

Query Match 33.5%; Score 490; DB 2; Length 1812;
Best Local Similarity 61.2%; Pred. No. 1.6e-140;
Matches 887; Conservative 0; Mismatches 515; Indels 48; Gaps 4;

Qy 33 CGAGTACAGAAATCCGAGATCAACCTCCATCTTCGCGCGTTCAGGAACTTACCCCTTC 92
Db 99 CGAGTACAGAAATCCGAGATCAACCTCCATCTTCGCGCGTTCAGGAACTTACCCCTTC 158
Qy 93 TCAGAGAAATCCGAGATCAACCTCCATCTTCGCGCGTTCAGGAACTTACCCCTTC 152
Db 159 CGAGTACAGAAATCCGAGATCAACCTCCATCTTCGCGCGTTCAGGAACTTACCCCTTC 218
Qy 153 GAGGAGTACAGAAATCCGAGATCAACCTCCATCTTCGCGCGTTCAGGAACTTACCCCTTC 212
Db 219 TAAGAGTACAGAAATCCGAGATCAACCTCCATCTTCGCGCGTTCAGGAACTTACCCCTTC 278
Qy 213 GACATCAAGAAATCCGAGATCAACCTCCATCTTCGCGCGTTCAGGAACTTACCCCTTC 272
Db 279 AACCTTACAGAAATCCGAGATCAACCTCCATCTTCGCGCGTTCAGGAACTTACCCCTTC 338
Qy 273 AGATACAGCGCTGCTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 332
Db 339 AGATACAGCGCTGCTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 371
Qy 333 ATCCAGTCTTCCGCTGAGAGGCGAAACACTCCAGAAATCTCCGAGAACTTACCCCTTC 392
Db 372 CGCGCGCTGCTGCTGCTGAGAGGCGAAACACTCCAGAAATCTCCGAGAACTTACCCCTTC 431
Qy 393 CGCTTCCATGCGAGATGCTCAAGGCGCCACAGAGGCTTTCGATGATGCTGCTGCTGCTGCT 452
Db 432 GGCATCTGCTGCGGCTCAAGGCTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 491
Qy 453 TACACTCTCATCTCCAAAGGCTTCGAAATTCGAAACAGCGGCTGCTGCTGCTGCTGCTGCTGCT 500
Db 492 TACACTCTCATCTCCAAAGGCTTCGAAATTCGAAACAGCGGCTGCTGCTGCTGCTGCTGCTGCT 551
Qy 501 TCCAGAGCGCAACAGAGCTGACAACTCCGATTAACCTGCGCTGCTGCTGCTGCTGCTGCTGCT 560
Db 552 CCGAGATCTTAACTTACCGATTAACCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 611
Qy 561 GGTCTTCAACCAAGACAGAACTGCGACAGAGTTCGCGCGCTGCTGCTGCTGCTGCTGCTGCT 620
Db 612 AAGTTGAGACTGATCTCTTAAATATACCAAGTGAAGAAAGACTGCTGCTGCTGCTGCTGCT 671
Qy 621 CGAAGAGCAACCAAGAGTGTCCACCGCTCTACCGCTTACAGAGGAGGAGGAGGAGGAGGAG 680
Db 672 TGAGGAACTTACCCTGAGTTAAGAGGCTTTATCAGATGCGAGGCTTAATGGAATCTTGTCT 731
```

QY 681 CTTCCAGCCATCAACCTCAAGCGCTGTTCACAAAGTCCAAAGTTCGATACATCACTACGG 740  
Db 732 TTTCCCTCTCTATTAATGTAATCTGTACCAAGCAAGTTCGACAACTGTACGG 791  
QY 741 CTGCGCCACATCCCTTATGATGATGATCAACCGTCTCGATGATCATGATCGGGGCAA 800  
Db 792 ATGCGCCACATCACTGCGCGATGGTCTCATGAGGCTACTGATGATGATGATGCGGAAA 851  
QY 801 GACAGCTCTCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 860  
Db 852 GGTTCGCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 911  
QY 861 CCAAGCGCTCGCTGTATCATCAAGACTGACCAATCTGCGCTCTCCAGGCTGCAT 920  
Db 912 AGCGCGCTCGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 971  
QY 921 GGAAGCTTACAGCTCGCGCATCGAGAGTGTCAAGAGTGTCAAGAGTGTCAAGAGTGTCA 980  
Db 972 GGAAGCTTACAGCTCGCGCATCGAGAGTGTCAAGAGTGTCAAGAGTGTCAAGAGTGTCA 1031  
QY 981 ATGACAGAAATGCGATATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1040  
Db 1032 CACGACCGGTAAAGGACATTAATCATGCTTGACCAATGAGGAGATGAGGAGCAATGC 1091  
QY 1041 TATTCTCGTAACTCGGCACTTCGATAAGAAATGATACAGATGGCTCTCAAGATA 1100  
Db 1092 CATTTTGGCAATTTGTCATCTTGACAAAGAAATGACATCTTGGTCTCGAGACCTA 1151  
QY 1101 CCGAGGCTCAAGCACATCCCAATCAAGCAAGATGACATGTGGAAATCCCAAGATGG 1160  
Db 1152 CCTGCTGTAAGAGGATCAATTAAGCTCAACCGACAGATGGTCTTCCCTGACAC 1211  
QY 1161 CCAC---GCTATCT 1217  
Db 1212 CAACAGTGGCATATCT 1271  
QY 1218 TCACCCATCTTCT 1277  
Db 1272 ACACCTAGTTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1331  
QY 1278 CT-----ACGAAAGAGAGAAATCTCGAAGAGAGTTTACACATCTTCCGAAAGCATCT 1331  
Db 1332 GTGGAATGAAG 1391  
QY 1332 CGATGAAG 1391  
Db 1392 CGACGAG 1451  
QY 1392 ACAG 1451  
Db 1452 GAG 1511  
QY 1452 CCGTTATTAA 1461  
Db 1512 CAGGTACTGA 1521

## RESULT 4

US-08-930-894-1  
; Sequence 1, Application US/08930894  
; Patent No. 6037524  
; GENERAL INFORMATION:  
; APPLICANT: GREENLAND, Andrew James  
; APPLICANT: DRAPER, John  
; APPLICANT: SKIPSEY, Marc  
; APPLICANT: WARNER, Simon  
; TITLE OF INVENTION: S-ADENOSYL-L-HOMOCYSTEIN HYDROLASE PROMOTER  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pillsbury Madison & Sutro  
; STREET: 1100 New York Avenue, N.W.  
; CITY: Washington

STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MS Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/930,894  
FILING DATE: 09-OCT-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB96/00882  
FILING DATE: 10-APR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9507381.3  
FILING DATE: 10-APR-1995  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1767 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: SHH GENE FROM ASPARAGUS  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 26..1483  
OTHER INFORMATION: /codon\_start= 26  
US-08-930-894-1

Query Match 33.5%; Score 489; DB 3; Length 1767;

Best Local Similarity 61.1%; Pred. No. 3.2e-140;  
Matches 886; Conservative 0; Mismatches 515; Indels 48; Gaps 4;

QY 33 CGAGTACAGAAATGCGGACATCAACCTCCATGTTCTGGCGGTAAGGAACATTACCCCTTGC 92  
Db 61 CGAGTACAGAAATGCGGACATCAACCTCCATGTTCTGGCGGCTCGAGATCGAGCTCGC 120  
QY 93 TGAGAGAAATGCGGAGCTTATGTTCTGTTGAGCGCTTATTCGCTTCTAAGCCATT 152  
Db 121 TGAGTGCAGATGCCAGGCTCATGGCTCGCGGCTGAGTTGGCGCCGCCAGCCATT 180  
QY 153 GAAGGGTGTGAGAAATCTCTGGTTCCCTCCATGACAGATCCAGAGAGCGCTCTCATCGA 212  
Db 181 CAAGGGCGCAAAATCACTGGATCCCTCCATGACAGATCCAAATGCGCTCTCATCGA 240  
QY 213 GACACTCAGAGCTTGT 272  
Db 241 AACCCCTAACCCCTCGGCGCGAGGTTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300  
QY 273 AGATACAGCGCTGCTGTATGTTGCGGCCAACAGGACACACAGAGAGCGCGCG 332  
Db 301 GGACCATGCGCGCTGCCATTGCCCGTACTCGC----- 336  
QY 333 TATCCAGTCTTTCCTTGAAGGGGAAACACTCCAGAAATCTGGGAGAAACATACCG 392  
Db 337 ---CTCCGCTTTCCTTGAAGGGTGAACCTCCAGGAGTACTGGTGTGTGTGTGTGTGTGT 393  
QY 393 CGCTCTCAGATGGCCAGATGTTCAAGGCCACAGAGGTTGTGTGTGTGTGTGTGTGTGTGT 452  
Db 394 TGCCTCTGACTGGGGGCCCGCGGCTGACCTCATGTCGATGACGCGCGCGCAC 453  
QY 453 TACACTCTCTCATCTCCAGGGCTT-----CGAATTCGAAACAGCGCGTGTGT 500  
Db 454 CACTCTCTTGTATCCATGAGGGGTGAAGCCGAGAGAGTACGAGAGAGCGGGAGAT 513  
QY 501 TCCAGAGCAACAGAGCTGACAACTCCGATACCGCTGCTTCTGTCTACATCAAGCA 560



Db 514 GCCCGATCGCGCTTACCGACAACTGCTGAGTTCCAGATCGTCTCACATCATCAGGA 573  
Qy 561 GGTCTTCAACCAAGAACCACTGGCAACAGTGTCTGCGCGCATGAAGGTGTTTC 620  
Db 574 TGGGCTCAAGGTGACCCACCAAGTACAGGAAGATGAAGATGATGTTGCGTGTGTC 633  
Qy 621 CGAAGAGACAAACAGGTGCTCAACCGCTCTACCGCTTCCAGCTGAGAGAGGGCAACCT 680  
Db 634 GGAGAGACCAACCGGGGTCAAGAGGCTTTACAGATCGAGCTAAACAATCCCTTCT 693  
Qy 681 CTTCCAGCCATCAACGCTCAACGACGCTTTACAAAGTCCAAAGTTCGATAACATCAGG 740  
Db 694 TTTCCCTGGATCAATGTCAATGACTCCGTCAACAGAGCAAGTTTGACAATCTGTAGG 753  
Qy 741 TTGCGCGCACTCCCTTATCGATGATATCAACCGTGTCTCCAGTGTGATGATCGCGGCA 800  
Db 754 ATGCGCGCACTCTCTTCCGATGCTGATGAGGCGCCACTGATGTTATGTTGCTGGCAA 813  
Qy 801 GACAGCTCTCGTCAATGGGTACGGGATGTGGCAAGGCTGGCTCAATCCCTCGTGG 860  
Db 814 GGTTCGAGTTGTCTGCGGTTATGGTGATGTGGAGAGGCTGTGCTGCTGACTCAGCA 873  
Qy 861 COAAGGCTCGCTTATCATCAAGAACTGAAACCACTGCGCTCTCCAGGTGCCAT 920  
Db 874 GGTGCTGCTGCTTATGCTGACGGAGATCGAACCATCTGTGCTCTTCAAGCCCTAAT 933  
Qy 921 GGAAGCTTACAGTCTCGCGCATCGAGGAGTGTCAAGGATGTGATATCTTCTGTTAC 980  
Db 934 GGAGGCTCTTCAAGTCTCCCTCGAGATGTTGTCTCAGAGGCGGATATCTTGTATC 993  
Qy 981 ATGCAAGAAATCGCGATATCATCTCTGTGATGATGCGCCAGATGAAGGATAAGGC 1040  
Db 994 CACCAACGGTAAACAGACATCATCATGCTGGACCATGAGGAAGTGAAGCAATGC 1053  
Qy 1041 TATTCTCGGTAACTCGGCACTTGATTAAGAAATGATACAGATGCGCTCATGAATA 1100  
Db 1054 CATCTCTGCAACATTTGCTCACTTTGCAACGAGATTGACATCTAGGTTTGGAGACATA 1113  
Qy 1101 CCCAGCATCAAGACATCCCAATCAAGCCAGATAGCATGTGGGATTTCCAGATGG 1160  
Db 1114 CCTTGGCATCAAGAAATCAACATCAAGCCCAAGCTGACCGGTGGGTCTTCCCTGAAC 1173  
Qy 1161 CCAC---GCTATCCTCTCTTCTGAGGCGCGCTCTTAACTTGGCTGCGCTACAGG 1217  
Db 1174 CAACACTGGTATATTTCTTCTGAGGCGGCTACTATGAACTTGGGTGGCACTGG 1233  
Qy 1218 TCACCCATCTTCTGTTATGTAATGTCATTCACAAACAGACACTCGCTCAGCTGACCT 1277  
Db 1234 TCACCCAGCTTTGTCTATGTCCTGCTCTTCAACCAACAGGTGATTTGCTCAGTAGTT 1293  
Qy 1278 CTACGAAAG-----AGAGAAATCTCGAGAGAGGTTTACACATTCGAGCATCT 1331  
Db 1294 GTGGAATGAGAGCAAGCGGCAAGTATGAGAGAGGTTTACGTCTCCCCAGCATCT 1353  
Qy 1332 CGATGAGAGAGTCTGCTCGCTCCACCTCGGATCTCTGATGTCGATTCGACCTTACAAAGCTTAC 1391  
Db 1354 TGATGAGAGAGTASCAGCGCTTCACTTGTGGCAAGCTGGAGCCAAAGCTTACAAAGCTCAG 1413  
Qy 1392 ACAGAGAGAGCTGACTACATCAAGCTTCCAGTGTGAGGTCCTTAAAGCTGTGATGCTTA 1451  
Db 1414 CCTTTCAAGCGGAGCTACATCAGCTGCCATGCCATGCCATGCCATGCCATGCCATGCCAT 1473  
Qy 1452 CGGTATTA 1460  
Db 1474 CAGGTACTA 1482

## RESULT 5

US-09-318-448-26  
; Sequence 26, Application US/09318448  
; Patent No. 6210950  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, William G.

; APPLICANT: Stenroos, Edward S.  
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING  
; FILE REFERENCE: 601-1-057  
; CURRENT APPLICATION NUMBER: US/09/318,448  
; CURRENT FILING DATE: 1999-05-25  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 26  
; LENGTH: 2211  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-318-448-26

Query Match 15.3%; Score 223.2; DB 3; Length 2211;  
Best Local Similarity 55.3%; Pred. No. 2.5e-58;  
Matches 553; Conservative 0; Mismatches 343; Indels 104; Gaps 2;  
Qy 563 TCTTCAACCAAGACAAAGACACTGGCACACAGTGTCTGCGCGCATGAACGCTTTTCG 622  
Db 451 TCACCAACCTCATCCACACCAAGTACCCGCGAGCTTTGCGCAGGCAATCCGAGGCACTCTG 510  
Qy 623 AAGAGACAAACAAAGGTGTCCACCGCTCTACAGAGTTCGAGAGAGGAGGCGCAAACTCCTCT 682  
Db 511 AGGAGACACAGACTGGGGTCCACACCTCTACAGATGATGGCCATGGGATCCTCAGG 570  
Qy 683 TCCAGCCATCAACGCTCAACGACGCTGTTCAAAGTCCCAAGTTCGATTAACATCTACCGCT 742  
Db 571 TGCCTGCCATCAATGTCAATGACTCCGTCCCAAGAGCAAGTTTCACAACCTCTATGGCT 630  
Qy 743 GCGGCCACTCCCTTATCGATGTATCAACGCTGTCTCCGATGTCTGATGATGATGATGATGAT 802  
Db 631 GCGGGAGTCCCTCTAGATGGCATCAAGCGGGCCACAGATGTGATGATGATGATGATGATGAT 690  
Qy 803 CAGCTCTCGTCTAGGGTTACGGCGATGTGCGCAAGGCTGCGCTCAATCCCTCCGTTGCC 862  
Db 691 TAGCGGTGGTAGCAGCTATGTGTATGTGGCAAGGCTGTGCCAGGCTCTGCGGGTT 750  
Qy 863 AAGGGCTCGGTTATCATCAGAACTCGAACCAATCTGCGCTCTCCAGGCTGCCATGG 922  
Db 751 TCGGAGCCCGCTCATCATCCGAGATTGACCCCATCAACGCACTGCGAGCTGCCATGG 810  
Qy 923 AAGGCTACAGGCTCCCGCATCGAGGAAGTCGTAAGGATGTGATATCTTCTGTTACAT 982  
Db 811 AGGCTATGAGGTGACCACTATGATGAGGCTGTGAGAGGCAACATCTTTGTACCA 870  
Qy 983 GCACAGAAACTCGCATATCATCTCTGTGACATGATG----- 1020  
Db 871 CCACAGGCTGTATTGACATCATCTTGGCGGTAGGTGCCAGATGGGGGTCCCGGGAG 930  
Qy 1021 ----- 1020  
Db 931 TGAGGAGGAGGACAGATTGGGACAGCTTTCTGTCCCTGACAAATCTCCACGCTTTGG 990  
Qy 1021 -----GCCACAGATGAAGATAAGGCTATTGTGGTAAACATCGGCCA 1061  
Db 991 GCTCCTGACAGGCACTTTGACGACATGAAGATGATGCCATTGTGTACATTTGACA 1050  
Qy 1062 CTTGATTAAGAAATGATACAGATGGCTCATGAATACCCAGGATCAAGCAATCCC 1121  
Db 1051 CTTTGAAGCTGAGATCGATGTCAAGTGGCTCAACGAGAAC---GCCGTGAGAGAGGTGAA 1107  
Qy 1122 AATCAAGCCAGATACAGATGTGGGATTTCCAGATGGCAGCTATCTCTTCTTGC 1181  
Db 1108 CATCAAGCCGAGGTGACCGGTATCGGTTGAAGATGGGCGCGCATCTCTGCTGGC 1167  
Qy 1182 TGAGGGCGGCTTTTAACTTGGCTGCGCTACAGGTTCACCATCTTTCTGTTATGTCAAT 1241  
Db 1168 CGAGGCTCGGCTGGTCAACTGGTTGTGCCATGGGCGCACCCAGCTTCGTGATGATTA 1227  
Qy 1242 GTCAATCAAAACAGACACTCGCTCAGCTCGACCTCTTACGAAAGAGAGGAAATCTCGA 1301  
Db 1228 CTCCTTCAACCAACAGGTGATGGCGCAGATCGAGCTGTGGACCCATCCAGACAAGTACC 1287

QY 1302 GAAGAAGGTTTACACATCTCCGAGCAGTCTGATGAAGAAAGTGGCTCGCTCCACCTGG 1361  
Db 1288 CGTGGGTTTCAATTTCTGCTCCCAAGAGCTGGATGAGCAGTGGCTGAAGCCACCTGGG 1347  
QY 1362 ATCTCTGATCTCCACCTTACAAAGCTTACACAGCAGGCTGACTATACAAAGTTC 1421  
Db 1348 CAACTGATGTGAAGTTGACCAAGCTAACTGAGAGCAAGCCACGACTACCTGGCATGTC 1407  
QY 1422 AGTTGAGGTCCTTACAAAGTCTGATGCTTTACCGTTATTAA 1461  
Db 1408 CTGTGATGCCCCCTTCAAGCCGGATCACTACCGTACTGA 1447

RESULT 6  
US-09-347-878-2  
; Sequence 2, Application US/09347878C  
; Patent No. 6376210  
; GENERAL INFORMATION:  
; APPLICANT: Yuan, Chong  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ASSAYING ANALYTES  
; FILE REFERENCE: 25885-1651  
; CURRENT APPLICATION NUMBER: US/09/347,878C  
; CURRENT FILING DATE: 1999-07-06  
; NUMBER OF SEQ ID NOS: 75  
; SOFTWARE: Patenlin Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 2211  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Human S-adenosylhomocysteine hydrolase cDNA  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: M61831/GenBank  
US-09-347-878-2

Query Match 15.3%; Score 223.2; DB 4; Length 2211;  
Best Local Similarity 55.3%; Pred. No. 2.5e-58;  
Matches 553; Conservative 0; Mismatches 343; Indels 104; Gaps 2;  
QY 563 TCTTCAACCAAGACAAGAACCACTGGCACACAGTTGCTGGCGGCATGAACGGTGTTC 622  
Db 451 TCACCAACTCATCCACACCAAGTACCAGCTCTGCGCAGGATCCGAGGCATCTCTG 510  
QY 623 AAGAGACACAGAGTGTCCAGGCTCTACAGCTCGAGAGGAGGCAACTCTCT 682  
Db 511 AGGAGCACAGACTGGGGTCCACAACTCTACAGATGATGGCAATGGGATCTCAAG 570  
QY 683 TCCAGCCATCAAGCTCAACGAGCTGTATACAAAGTCCAAAGTTCGATAACATCTACGGCT 742  
Db 571 TGCTTGCATCAATGTCAATGATCTCGTCCACCAAGAGCAAGTTTGACACCTCTATGCT 630  
QY 743 GCGGCATCTCCCTATCGATGGTATCAACCGTGTTCGATGTCATGATCGCGGCAAGA 802  
Db 631 GCGGGAGTCCCTCATATAGATGGATCAAGCGGGCCACAGATGTGATGATTCGCGCAAG 690  
QY 803 CAGCTCTGTCTATGGGTATCGGATGTCGCAAGGGCTGGCTCAATCCCTCCGTGCC 862  
Db 691 TAGCGTGTGAGAGGCTATGGGTATGGGCAAGGGCTGGCCAGGCTTCGCGGGTT 750  
QY 863 AAGCGCTCTGGGTATCATACAGAACTCGACCCCAATCTGGCTCTCCAGGCTGCCATGG 922  
Db 751 TCGAGCCCGGTCTATCATCCAGAGATGACCCCATCAAGCACTCGAGGCTGCCATGG 810  
QY 923 AAGCTTACAGGTTCGCGGATCGAGGAAGTGTCTCAAGATGTGATATCTTCTGTTACAT 982  
Db 811 AGGCTTATGAGTGAACCACTGATGAGGCTGTGAGGAGGCAACATCTTTGTCAACA 870  
QY 983 GCACAGAACTGGGATATCATCTCTGTTGATGATG----- 1020  
Db 871 CCACGGCTGTATTGACATCATCTTGGCCGGTAGGTGCCAGATGGGGGGTCCCGGGAG 930  
QY 1021 ----- 1020

Db 931 TGAGGAGGAGGAGCAGAGTTGGGACAGCTTTCTGTCCCTGACAAATCTCCACGGTCTGG 990  
QY 1021 -----GCCAGATGAAGGATAAGGCTATTGTGCGTAACATCGGCCA 1061  
Db 991 GTCCTCGACAGGCACCTTTGACGAGATGAAGGATGATGCTTGTGTAACTTGACA 1050  
QY 1062 CTTGATAAGAAATGATACAGATGGCTTCATGAATACCCAGGCATCAAGCAATCC 1121  
Db 1051 CTTTGAAGTGGAGATCGATGTCAAGTGGCTCAACGAGAAC---GCCGTGGAGAAGTGAA 1107  
QY 1122 AATCAAGCCAGATACAGATGTGGAAATCCAGATGGCCAGCTATCTCTCTTTCTTC 1181  
Db 1108 CATCAAGCCGAGGTGACCGGTATCGGTGAAGATGGCGCGCATCTCTCTCTGGC 1167  
QY 1182 TGAGGGCGGCTTCTTAACCTTGGCTGAGCTACAGGTCACCCATCTTTCTGTATGTCAAT 1241  
Db 1168 CGAGGCTCGCTGCTCAACTGGTGTGGCATGGGCCACCCACGCTTCGTGATGAGTAA 1227  
QY 1242 GTCTATTCAAAACAGACACTCTGCTGAGCTGACCTCTTACGAAAGAGAGGAATCTCGA 1301  
Db 1228 CTCCTTCAACCAAGGTGATGGCGCAGATCGAGCTGTGACCCCATCCAGACAAGTACC 1287  
QY 1302 GAAGAAGGTTTACACACTTCGAAAGCATCTCGATGAAGAGTGGCTCGCTCCACCTCG 1361  
Db 1288 CGTGGGTTCTTCTGCCCCAGAGCTGAGTGGAGGAGTGGCTGAAGCCCACTGGG 1347  
QY 1362 ATCTCTGATGTCCACTTACAAAGCTTTACAGAGCAGGCTGACTATACAAAGTTC 1421  
Db 1348 CAAAGTGAATGTGAAGTGTGACCAAGCTAACTGAGAAGCAAGCCAGTACCTGGCATGTC 1407  
QY 1422 AGTTGAGGTCCTTACAAAGTCTGATGCTTACCGTTATTAA 1461  
Db 1408 CTGTGATGGCCCCCTTCAAGCCGGATCACTACCGTACTGA 1447

## RESULT 7

US-09-252-991A-6815/c  
; Sequence 6815, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6815  
; LENGTH: 1485  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6815

Query Match 12.3%; Score 179.4; DB 4; Length 1485;  
Best Local Similarity 54.3%; Pred. No. 6e-45;  
Matches 511; Conservative 0; Mismatches 346; Indels 84; Gaps 4;  
QY 605 GCATGAACGGTGTTCGAGAGACAAACAGGTGTCCACCGCTCTACAGCTCAGA 664  
Db 1010 GCATCCACGGCATCACCGAAGAGACCAACCGGGGTCCACCGCTGTGCATGTCTCA 951  
QY 665 AGGAGGCAAACTCCTTCTCCAGCCATCAACGTCACACGCTGTACAAAGTCCAAAGT 724  
Db 950 AGAAGCGCGCTTGAAGTCCCGGGATCAACGCTCAACGCTCGGTGACCAAGCAGA 891  
QY 725 TCGATACATCTAGGCTCGCCCACTCCCTTATCGATGTATCAACCGTCTTCCGATG 784  
Db 890 ACGACAAAGTACGGCTGCGCCATAGCTTCAACGAGCCATCAAGCGCGGCGCACGCC 831

785 TCATGATCGGCGGCAAGACAGCTCTCGTCATGGGTACGGCGAGTGCAGAAAGGCTGCG 844  
Db ACCTGCTGTCGGCGAAGCAGGCGCTGCTGATCGGTACGGCGAGCTGGGCAAGGCGCTCCT 771  
845 CTCATCTCCTCGGTGGCCAGGCGCTCGGGTATCATCAGAACTCGAACCCCAATCTGCG 904  
Db CGCATGCTCGGCGGAGAGGATGATCGTCAAGGTGCGGGAAGTGCAGCCGATCTGCG 711  
905 CTCCTCAGGCTGCCATGGAAGGCTACAGGTC-----CGCCGATCGAGG 949  
Db CCATCGAGGCTGTCATGAGCGCTTCGAGTGTCTCGCGTCAAGAAAGCGCATCAAG 651  
950 AAGTGTCTAAG-----ATGTCGATATCTTCGTACAT 982  
Db AGGCAACCGAGGCGAGCATGACGCGCGCTGCTGGCAAGATCGACCTGATCGTCAACA 591  
983 GCACAGGAACTGCGATATCATCTCTGTTGACATGATGGCCAGATGAAGGATAGGCTA 1042  
Db CCACCGGCAAGTCAAGTCTGCGAGCGCCAACTGCTCAAGGCGCTGAAGAGCGCGCG 531  
1043 TTGTCGTAATCATCGGCGCTTCGATPACGAAATGATACAGATGCGCTCATGAATATACC 1102  
Db TGGTCTGCAATCATCGGCGCTTCGACACAGATCGACCGGCTTCATGCGCAAGAACT 471  
1103 CAGGCAATCAA-----GCACATCCCAATCAAGCCAGATACGACATGT 1144  
Db GGGCTTGGGAAGAGGTCAAGCGCAGGTGCAAGATCCACCGCACCGCAAGAGCGTT 411  
1145 GGGAAATCCAGATGCGCAGCTATCTCTCTTGTGTGAGGCGCGCTTCTTAACCTTG 1204  
Db TCAGCGGCAACAGCATCTACTGATCTCTCTCGCGAGGCGCGCTTGTCAACCTGG 351  
1205 GTCGCGTACAGGTACCCATCTTCTGTTATGTCATGTCATTCACAAACAGACACTCG 1264  
Db GCAAGCGCACCGGCGCTATCGAGCGGATCATGACCGTCTCTTCGCCAACAGGAGTGTG 291  
1265 CTCAGTCGACTCTAGAAAGAGAGGAAATCTCGAA-----1304  
Db CGAGATCACTCTGTCAGACAGTACGCCGACCTCGCGCGCGCGGAGAGGCCAAGC 231  
1305 ----GAAGGTTTACACACTTCCGAAGCATCTCGATGAAGAGTGTGCTGCCCTCCACTCG 1360  
Db GCTGAGGCTGGAAGTCTCCGAAGACTCGACGAGGAGTGGCCCTGGAGATGCTCA 171  
1361 GATCTCTGATGTCACCTTACAAAGCTTACAGAGAGCGGCTGACTCATCAACGTTTC 1420  
Db AGGGCTTTCGGCGGCTGTGTCACCCAACTGACCCGAAACAGCGCGAGTACATCGCGCTGA 111  
1421 CAGTTGAGGTCCTTACAGTCTGATCGTTCAGGTTATTAA 1461  
Db GGTGGAAGGCGGCTTCAAGCGGACACCTATCGCTACTAA 70

RESULT 8  
US-09-252-991A-6930  
; Sequence 6930, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubinfeld et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6930  
; LENGTH: 1509  
; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6930  
Query Match 12.3%; Score 179.4; DB 4; Length 1509;  
Best Local Similarity 54.3%; Pred. No. 6.1e-45;  
Matches 511; Conservative 0; Mismatches 346; Indels 84; Gaps 4;  
Qy 605 GCATGAACGGTGTTCGGAAGAGACAACAAGGTGTCCACCGCTCTTACAGTCCAGAG 664  
Db GCATCCACGGCATCACCGAAGAGACCAACCGGGGTCCACCGCTGTCTGCATGCTCA 628  
Qy 665 AGGAGGCAAACTCCTTCCAGGCNTCAAGCTCAACGACGCTGTACAAGTCCAGT 724  
Db AGACGGCGCCCTGAAGTCCCGGCGATCAAGCTCAACGACTCGGTGACCAAGAGCAGA 688  
Qy 725 TCGATAACATCTACGGCTCCGCCACTCCTTATCGATGTATCAACCGTGTCTCCGATG 784  
Db ACGACAACAAAGTACGGCTCCGCCATAGCTCAACGAGCGCATCAAGCGCGCACCGACC 748  
Qy 785 TCATGATCGGCGGCAAGACAGCTCTCGTCATGGTGTACGGCGATGTGCGCAAGGCTGCG 844  
Db ACCTGCTGTGCGGCAAGCAGCGCCCTGCTGATCGGCTACGGCGACGTGGCAAGGCTCCT 808  
Qy 845 CTCATATCCTCCGTGCGCAAGCGCTCGCGTTATCATCAAGAACTTCGACCCCAATCTGCG 904  
Db CGCATGCTGCGCGCAAGAGGAGCATGATGTCAGGTGCGGAGTGCAGACCGGATCTGCG 868  
Qy 905 TTCTCCAGGCTGCCATGGAAGCTTACAGGTC-----CGCCGATTCGAGG 949  
Db CCATCGAGGCTGCGATGAGCGGCTTCGAAGTCTGTCGCGTACAAAGACGSCATCAAG 928  
Qy 950 AAGTGTCAAG-----ATGTCGATATCTTCGTACAT 982  
Db AGGCAACCGAGGCGCAGCATCGACCGCGCTGCTGGGCAAGATCGACCTGATCGTCAACA 988  
Qy 983 GCACAGGAACTCGATATCATCTCTGTGACATGAGCCCGAGATGAAGGTAAGGCTA 1042  
Db CCACCGGCAACGTCACGCTCTCGGAGCGCAACATGCTCAAGGCGCTGAAGAGCGCGGG 1048  
Qy 1043 TTGTCGTAATCATCGGCGCTTCGATTAAGAAATGTATACAGATGCGCTCATGAATAACC 1102  
Db TGGTCTGCAATCATCGGCGCTTCGACACAGAGATCGACACCGCTTCATGCGCAAGAACT 1108  
Qy 1103 CAGGCAATCAA-----GCACATCCCAATCAAGCCAGATACGACATGT 1144  
Db GGGCTTGGGAAGAGGTCAAGCGCGAGGTGCAAGATCCACCGCACCGGCAAGGACGCTT 1168  
Qy 1145 GGGAAATCCAGATGCGCACGCTATCTCTTCTGTCGAGGCGCGCTTCTTAACCTTG 1204  
Db TCAGCGCGCAACAGCGACTACTCTGATCTCTCGCGAAGCGCGCTGCTCAACCTGG 1228  
Qy 1205 GCTGCGCTACAGGTCAACCATCTTTCTGTTATGTCATTCATTCACAAACAGACACTCG 1264  
Db GCAAGCGCACCGGCGCATCCGAGCGGATCATGACGCTTCTTCGCCAACAGGAGTGTG 1288  
Qy 1265 CTCAGCTCGACTCTAGAAAGAGAGGAAATCTCGAGAA-----1304  
Db CGCATATCCACTGTTCGAGAGAGAGTACGCCGACCTGCGCGCGCGGAGAGGCCAAGC 1348  
Qy 1305 ----GAAGGTTTACACACTTCCGAAGCATCTCGATGAAGAGTGTGCTGCCCTCCACCTCG 1360  
Db GCTGAGGCTCGAAGTGTCTCGCGAAGAGCTTCGAGGAGGAGTGGCCCTGGAGATGCTCA 1408  
Qy 1361 GATCTCTGATGTCACCTTACAAAGCTTACAGAGAGCGGCTGACTCATCAACGTTTC 1420  
Db AGGGCTTTCGGCGGCTGTGTCACCCAACTGACCCGAAACAGCGCGAGTACATCGCGCTGA 1468  
Qy 1421 CAGTTGAGGTCCTTACAGTCTGATCGTTCAGGTTATTAA 1461  
Db GGTGGAAGGCGGCTTCAAGCGGACACCTATCGCTACTAA 1509

RESULT 9

Qy	1307	AGGTTACACACTTCGGAAGCACTCGATGAAGAAGTCGCTCGCTCCACTCGGATCTC	1366
Db	1238	CGGTAGAAGTTCTTCCTAAGAACTTGACGAAGAGGTTTCCTGTCGAATGGTTGCAGGTT	1297
Qy	1367	TCGATGTCACACTTACAAGCTTACACAGAAGCAGGCTGACTATCATCAAGCTTCCAGTTG	1456
Db	1298	TCGGTGTGTACTTACTCAGTTAACTCAAGACCAAGCTGATTTATTAGCATTTGCTGTTG	1357
Qy	1427	AGGTCCTTACAAGTCTCGATGCTTACCGTTATTAA	1461
Db	1358	AGGGCCCAATCAAACTCTGACGCTTACAAATACTAA	1392
RESULT 10			
US-09-702-705-1179			
; Sequence 1179, Application US/09702705			
; Patent No. 6504010			
; GENERAL INFORMATION:			
; APPLICANT: Wang, Tongtong			
; APPLICANT: Bangur, Chaitanya S.			
; APPLICANT: Lodes, Michael A.			
; APPLICANT: Fanger, Gary			
; APPLICANT: Vegdick, Tom			
; APPLICANT: Carter, Darick			
; APPLICANT: Retter, Marc			
; APPLICANT: Mannion, Jane			
; APPLICANT: Fan, Liqun			
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND			
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER			
; FILE REFERENCE: 210121.478C14			
; CURRENT APPLICATION NUMBER: US/09/702,705			
; CURRENT FILING DATE: 2000-10-30			
; NUMBER OF SEQ ID NOS: 1833			
; SOFTWARE: FastSeq for Windows Version 3.0			
; SEQ ID NO 1179			
; LENGTH: 288			
; TYPE: DNA			
; ORGANISM: Homo sapien			
US-09-702-705-1179			
Query Match 10.5%; Score 153.8; DB 4; Length 288;			
Best Local Similarity 71.2%; Pred. No. 1.6e-37;			
Matches 203; Conservative 0; Mismatches 82; Indels 0; Gaps 0;			
Qy	668	AGGCAAACTCTCTTCCCAACCATCAAGCTCAACGACGCTGTTACAAAGTCCAAAGTTTCG	727
Db	4	ATGGATCTCTCAAGGTGCTGCCATCAATGTCAATGACTCCGTCAACAGAGCAAGTTG	63
Qy	728	ATACATCTACGGCTGGCCCACTCCCTTATCGATGGTATCAACGCTGCTTCGATGTCA	787
Db	64	ACAACCTCTATGCTGCGGAGTCCCTCATAGATGGCATCAAGCGGCCACAGATGTGA	123
Qy	788	TGATCGCGCGCAAGACAGCTCTCGTCAATGGTTAGCGGATGTGGCAAGGGCTGCGCTC	847
Db	124	TGATTGCGGCAAGGTAGCGGTGTGTAGCGGTATGTGTATGTGGCAAGGGCTGTGCC	183
Qy	848	AATCCCTCGGTGCCAAGGCGCTCGGTTATCATACAGAACTCACCCAACTTGGGCTC	907
Db	184	AGGCCCTCTGGGGGTTTCGGAGCCCGGCTATCATCACCGAGGTTGACCCCATCAAGCAC	243
Qy	908	TCCAGGCTGCATGAAGGCTTACCAGGTCCGCGCATCGAGGAAG	952
Db	244	TGCAGGCTGCCATGAGGGCTATGAGGTGACCACCATGATGAGG	288
RESULT 11			
US-09-736-457-1179			
; Sequence 1179, Application US/09736457			
; Patent No. 6509448			
; GENERAL INFORMATION:			
; APPLICANT: Wang, Tongtong			
; APPLICANT: Bangur, Chaitanya S.			
; APPLICANT: Lodes, Michael A.			

APPLICANT: Fanger, Gary  
APPLICANT: Vedvick, Tom  
APPLICANT: Carter, Darrick  
APPLICANT: Retter, Marc  
APPLICANT: Mannion, Jane  
APPLICANT: Fan, Liqun  
APPLICANT: Wang, Aljun  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER  
FILE REFERENCE: 210121.478C15  
CURRENT APPLICATION NUMBER: US/09/736,457  
CURRENT FILING DATE: 2000-12-13  
NUMBER OF SEQ ID NOS: 1864  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1179  
LENGTH: 288  
TYPE: DNA  
ORGANISM: Homo sapien  
US-09-736-457-1179

Query Match 10.5%; Score 153.8; DB 4; Length 288;  
Best Local Similarity 71.2%; Pred. No. 1.6e-37;  
Matches 203; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 668 AGGGCAAACTCTCTCCAGCCATCAACGTCAACGACGGCTGTACAAAAGTCCAAAGTTGC 727  
DB 4 ATGGATCTCAAGTCCCTGCAATCATGTCATGCTCGTCAACAGACAAAGTTG 63  
QY 728 ATACATCTAGGTCGGCCACTCCCTATGATGATGATCAACCGTGTTCGGATGTC 787  
DB 64 ACAACCTCTATGCTGGCGGAGTCCCTCATAGATGGCATCAAGCGGGCCACAGATGTA 123  
QY 788 TGATCGGGGCAACAGACTCTGTCATGCTGATGAGGCTAGCGGATGTCGCAAGGCTCGGTC 847  
DB 124 TGATTCGGGCAAGTAGCGGTGAGAGGCTATGATGATGTCGCAAGGCTGTGCCC 183  
QY 848 AATCCTCCGTGGCAAGGCGCTCGCTTATCATCAAGAACTCGACCAACATCTCGGTC 907  
DB 184 AGGCCCTCGGGGTTTCGGAGCCCGGTCATCATCACCGAGTTGACCCCATCAACGCAC 243  
QY 908 TCCAGGTGCATGGAAGGTACCGGTCGCGCCGATCGAGGAAG 952  
DB 244 TCCAGGCTCCATGGAGGCTATGAGGTGACCAACCATGGATGAG 288

RESULT 12  
US-08-204-740-8  
Sequence 8, Application US/08204740  
Patent No. 5753432  
GENERAL INFORMATION:  
APPLICANT: Gudkov, Andrei  
APPLICANT: Kazarov, Alexander  
APPLICANT: Mazo, Ilya  
APPLICANT: Roninson, Igor B  
TITLE OF INVENTION: Methods for Identifying Genetic  
TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant  
TITLE OF INVENTION: Growth in Cancer Cells  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Allgretti & Witcoff, Ltd.  
STREET: 10 S. Wacker Drive, Suite 3000  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/204,740  
FILING DATE: 04-MAR-1994

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5753432nan, Kevin E  
REGISTRATION NUMBER: 35,303  
REFERENCE/DOCKET NUMBER: 93,354-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
TELEX: 910-221-5317  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 289 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
US-08-204-740-8

Query Match 8.4%; Score 122.4; DB 1; Length 289;  
Best Local Similarity 71.1%; Pred. No. 7.6e-28;  
Matches 162; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 37 TACGAATGCGGACATCAACCTCCATGTTCTCGGCGTAAGGAACCTTACCTTGTCTGAG 96  
DB 62 TACAAGTCCGCGACATCGGCTTGGCTGGGAGCGAGGCCCTGGACATGCTGAG 121  
QY 97 AGGAAATGCGAGGTCTTATGTTTCTGTAAGCGTTTATTCGGCTTCTAAGCCATTTGAAG 156  
DB 122 AACGAGATGCGGCGCTGATGCGTATGCGGAGCGGTACTCGGCTTCCAAAGCCACTGAAG 181  
QY 157 GGTGTCAGATCTCTGTTCCCTCCATGACATGACATGACATGACATGACATGACATGACATG 216  
DB 182 GCGCGCCGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 241  
QY 217 CTCAGATCTTGTGCTGATGTCAGATGCGGCTTCTGCAACATCTTC 264  
DB 242 CTCGTCACCTGGTGTGCTGAGGTGCTGAGGTGCTGAGGTGCTGAGGTGCTGAGGTGCTGAGGT 289

RESULT 13  
US-09-081-167A-8  
Sequence 8, Application US/09081167A  
Patent No. 6083745  
GENERAL INFORMATION:  
APPLICANT: Gudkov, Andrei  
APPLICANT: Kazarov, Alexander  
APPLICANT: Mazo, Ilya  
APPLICANT: Roninson, Igor B  
TITLE OF INVENTION: Methods for Identifying Genetic  
TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant  
TITLE OF INVENTION: Growth in Cancer Cells  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
STREET: 300 S. Wacker Drive, 32nd Floor  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/081,167A  
FILING DATE: 18-MAY-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6083745nan, Kevin E  
REGISTRATION NUMBER: 35,303  
REFERENCE/DOCKET NUMBER: 93,354-KK  
TELECOMMUNICATION INFORMATION:

```
;
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-081-167A-8
Query Match 8.4%; Score 122.4; DB 3; Length 289;
Best Local Similarity 71.1%; Pred. No. 7.6e-28;
Matches 162; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
QY 37 TACAGATTGCCGACATCAACCTCCATGTTCTCGGCGGTAAAGAACTTACCCCTTCTGAG 96
Db 62 TACAAGTCGCGACATCGCCTGCTCGGGAGCGAGCCCTGGACATTCCTGAG 121
QY 97 AAGGAAATGCCAGGTCTTAAGTTCCTCGTAGCGTTATTCGGCTTCTAAGCCATTGAAG 156
Db 122 AACGAGATGCCGGGCGCTGATGCGGTATCGGGAGCGGTACTCGGCGCTCCAAGCCACTGAAG 181
QY 157 GGTGTCAAGATCTCTGTTCCCTCCACATGACATGCCAGACAGCGCTCTCATCGAGACA 216
Db 182 GCGGCCCGCATCGCTGCGCTGCCTGCACATGACCGTGGAGACGGCGCTCTCATGAGACC 241
QY 217 CTCACAGCTCTTGCTGCTGATGTCAGATGGGCTTCTCGCAACATCTTC 264
Db 242 CTCGTCAACCTGGGTGCTGAGTGCAGTGGTCCAGCTGCAACATCTTC 289
RESULT 15
US-09-416-833-8
; Sequence 8, Application US/09416833
; Patent No. 6197521
; GENERAL INFORMATION:
; APPLICANT: Gudkov, Andrei
; APPLICANT: Kazarov, Alexander
; APPLICANT: Mazo, Ilya
; APPLICANT: Roninson, Igor B
; TITLE OF INVENTION: Methods for Identifying Genetic
; TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant
; TITLE OF INVENTION: Growth in Cancer Cells
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allgretti & Witcoff, Ltd.
; STREET: 10 S. Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/416,833
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/204,740
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6197521nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,354-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 base pairs
```

```

; MOLECULE TYPE: cdna
US-09-416-833-8

Query Match      8.4%; Score 122.4; DB 3; Length 289;
Best Local Similarity 71.1%; Pred.No. 7.6e-28;
Matches 162; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 37 TACAGAAATGCGGACATCACCTCCATGTTCTCGGCGGTAAAGAACTTACCCTTCTCTGAG 96
Db 62 TACAAAGTCGCGGACATCGGCCCTGGCTGCTGGGAGCGCAAGGCCCTGGACATTGCTGAG 121

Qy 97 AAGGAAATGCCAGGTCTTATGTTTCTTCGTGAGCGTTATTCGGCTTCTAAGCCATTGAAG 156
Db 122 AAGGAAATGCCAGGTCTTATGTTTCTTCGTGAGCGTTATTCGGCTTCTAAGCCATTGAAG 181

Qy 157 GGTGTCAGAAATCTCTGGTTCCCTCCACATGACAGTCCAGACAGCCGTCCTCATCGAGACA 216
Db 182 GGTGTCAGAAATCTCTGGTTCCCTCCACATGACAGTCCAGACAGCCGTCCTCATCGAGACC 241

Qy 217 CTCACAGCTCTTGGTGTGCTGATGTCAGATGGGCTTCTTCGACACATCTTC 264
Db 242 CTCACAGCTCTTGGTGTGCTGATGTCAGATGGGCTTCTTCGACACATCTTC 289
```

Search completed: October 17, 2003, 22:56:13  
Job time : 139 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2003, 22:53:56 ; Search time 419 Seconds  
(without alignments)

9156.976 Million cell updates/sec

Title: US-09-759-990A-1

Perfect score: 1461  
Sequence: 1 atgcttgcaatacaactac.....ctgatgtctaccgtttattaa 1461

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1750203 seqs, 1313063994 residues

Total number of hits satisfying chosen parameters: 3500406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/1/pubna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/1/pubna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubna/US10\_NEW\_PUB.seq.\*  
16: /cgn2\_6/ptodata/1/pubna/US60\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/1/pubna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1461	100.0	1461	10	US-09-759-990-1
2	642.4	44.0	9025608	14	US-10-156-761-1
3	641.6	43.9	1455	14	US-10-156-761-5037
4	596.4	40.8	1557	11	US-09-746-660A-97
5	596.4	40.8	1939	11	US-09-919-854-1
6	596.4	40.8	3309400	10	US-09-738-626-1
7	594	40.7	1422	10	US-09-738-626-836
8	558.4	38.2	135638	12	US-10-314-657-1
9	529.6	36.2	1396	11	US-09-746-660A-101
10	331	22.7	2200	9	US-09-925-301-217
11	331	22.7	2429	13	US-10-044-090-344
12	302	20.7	3830	13	US-10-037-598-26
13	302	20.7	513509	12	US-09-754-853A-4
14	293.4	20.1	1398	12	US-10-342-224-51
15	288.4	19.7	720	10	US-09-738-626-838
16	278.6	19.1	708	10	US-09-738-626-837

17	244	16.7	2563	9	US-09-782-051-1	Sequence 1, Appli
18	238.4	16.3	874	13	US-10-027-632-151874	Sequence 151874, A
19	238.4	16.3	874	13	US-10-027-632-151875	Sequence 151875, A
20	223.2	15.3	2211	14	US-10-171-581-119	Sequence 119, App
21	186	12.7	448	11	US-09-918-995-9307	Sequence 9307, App
22	180.2	12.3	835	9	US-09-770-445-679	Sequence 679, App
23	177.2	12.1	963	13	US-10-027-632-153678	Sequence 153678, A
24	177.2	12.1	963	13	US-10-027-632-153679	Sequence 153679, A
25	177.2	12.1	963	13	US-10-027-632-153680	Sequence 153680, A
26	177.2	12.1	963	13	US-10-027-632-153681	Sequence 153681, A
27	171.8	11.8	424	10	US-09-960-352-9160	Sequence 9160, App
28	153.8	10.5	288	10	US-09-736-457-1179	Sequence 1179, App
29	153.8	10.5	288	10	US-09-902-941-1179	Sequence 1179, App
30	153.8	10.5	288	10	US-09-849-626-1179	Sequence 1179, App
31	153.8	10.5	288	12	US-10-113-872-1179	Sequence 1179, App
32	153.8	10.5	288	14	US-10-017-754-1179	Sequence 1179, App
33	153.4	10.5	284	10	US-09-920-300A-1423	Sequence 1423, App
34	153.4	10.5	284	12	US-10-099-326-1423	Sequence 1423, App
35	153.4	10.5	284	13	US-10-033-528-1423	Sequence 1423, App
36	151.8	10.4	288	10	US-09-998-598-2191	Sequence 2191, A
37	146.6	10.0	448	11	US-09-918-995-15324	Sequence 15324, A
38	146.2	10.0	728	9	US-09-770-149-117	Sequence 117, App
39	145.6	10.0	283	10	US-09-960-352-7006	Sequence 7006, App
40	137	9.4	793	10	US-09-966-881-9	Sequence 9, Appli
41	135.4	9.3	352	10	US-09-960-352-12928	Sequence 12928, A
42	133.2	9.1	410	10	US-09-960-352-1365	Sequence 1365, App
43	128.4	8.8	482	11	US-09-918-995-33549	Sequence 33549, A
44	124.2	8.5	389	10	US-09-960-352-10073	Sequence 10073, A
45	124	8.5	385	14	US-10-060-036-1769	Sequence 1769, App

#### ALIGNMENTS

#### RESULT 1

US-09-759-990-1  
; Sequence 1, Application US/09759990  
; Patent No. US20020119491A1  
; GENERAL INFORMATION:  
; APPLICANT: AntiCancer, Inc.  
; APPLICANT: Xu, Mingxu  
; APPLICANT: Han, Qinghong  
; TITLE OF INVENTION: HIGH EXPRESSION AND PRODUCTION OF HIGH  
; TITLE OF INVENTION: SPECIFIC ACTIVITY RECOMBINANT S-ADENOSYLHOMOCYSTEINASE  
; FILE REFERENCE: (SAHH) AND IMPROVED ASSAYS FOR S-ADENOSYLMETHIONINE (SAM)  
; CURRENT APPLICATION NUMBER: US/09/759,990  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: US 60/176,444  
; PRIOR FILING DATE: 2000-01-14  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1461  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Nucleotide sequence encoding SAHH  
US-09-759-990-1

Query Match	100.0%	Score 1461;	DB 10;	Length 1461;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1461;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ATGGCTTCAAAATCACCTACTGGTCTCCATTCGAGTACAGAAATGCCGATCAACCTC	60	
Db	1	ATGGCTTCAAAATCACCTACTGGTCTCCATTCGAGTACAGAAATGCCGATCAACCTC	60	
Qy	61	CATGTTCTCGCCGCTAAGGAACCTTACCTTCTCAGAGGAAGAAATGCCAGGCTTATGGTT	120	
Db	61	CATGTTCTCGCCGCTAAGGAACCTTACCTTCTCAGAGGAAGAAATGCCAGGCTTATGGTT	120	
Qy	121	CTTCGTGAGGCTTATTCGCGTCTTAAGCCATTGAAGGGTCTCAGAAATCTCTGTTCCCTC	180	



Db 121 CTTCTGAGCGTTATTCGCGTTCTAAGCCATTGAGGGTGTCAAGATCTCTGGTCCCTC 180  
QY 181 CACATGACATGTCAGACAGCGGTCCTCATCGAGACATCAAGCTCTTGCTGCTGATGC 240  
Db 181 CACATGACATGTCAGACAGCGGTCCTCATCGAGACATCAAGCTCTTGCTGCTGATGC 240  
QY 241 AGATGGGCTTCTCGACATCTTCTTCTACAGAGTACAGCGGCTGCTGCTGCTGCTG 300  
Db 241 AGATGGGCTTCTCGACATCTTCTTCTACAGAGTACAGCGGCTGCTGCTGCTGCTG 300  
QY 301 GGCCTCAAGAGGACACAGAGAGAGCGGCTATCCAGTCTTCGCTGGAAGGCGGAA 360  
Db 301 GGCCTCAAGAGGACACAGAGAGAGCGGCTATCCAGTCTTCGCTGGAAGGCGGAA 360  
QY 361 ACATCCCAAGATATCTGGAGAGACATACCGCGCTCTCAATGGCCAGATGGTCAAGGC 420  
Db 361 ACATCCCAAGATATCTGGAGAGACATACCGCGCTCTCAATGGCCAGATGGTCAAGGC 420  
QY 421 CCACAGCAGGTGTGATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 480  
Db 421 CCACAGCAGGTGTGATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 480  
QY 481 TTGCAAGAGCGGCTGCTGCTTCCAGAGCCAAACAGAGGCTGACAACTCGAATACCGCTGC 540  
Db 481 TTGCAAGAGCGGCTGCTGCTTCCAGAGCCAAACAGAGGCTGACAACTCGAATACCGCTGC 540  
QY 541 GTTCTTGCTACATCAAGAGGCTTCCAAACAGAGGCTGACAACTCGAATACCGCTGC 600  
Db 541 GTTCTTGCTACATCAAGAGGCTTCCAAACAGAGGCTGACAACTCGAATACCGCTGC 600  
QY 601 GCGCGCATGAAGCGGTGTTCCGAGAGAGACAAACAGAGTGTCCACCGCTCTACAGCTC 660  
Db 601 GCGCGCATGAAGCGGTGTTCCGAGAGAGACAAACAGAGTGTCCACCGCTCTACAGCTC 660  
QY 661 GAGAGAGAGGGAACCTCTTCCAGAGGCTGACAACTCGAATACCGCTGC 720  
Db 661 GAGAGAGAGGGAACCTCTTCCAGAGGCTGACAACTCGAATACCGCTGC 720  
QY 721 AGTTGCTGATACATCTAGGCTGCGCGCTCCCTTATPCGATGATATCAACCGTCTTCC 780  
Db 721 AGTTGCTGATACATCTAGGCTGCGCGCTCCCTTATPCGATGATATCAACCGTCTTCC 780  
QY 781 GATGTCATGATGCGGCGCAAGAGCTCTGTCATGAGGTTACGCGATGTGCGGAGGCG 840  
Db 781 GATGTCATGATGCGGCGCAAGAGCTCTGTCATGAGGTTACGCGATGTGCGGAGGCG 840  
QY 841 TCGGCTCAATCCCTCCGTCGCAAGCGCTCGCGTTATCATCAAGAACTCGACCAATC 900  
Db 841 TCGGCTCAATCCCTCCGTCGCAAGCGCTCGCGTTATCATCAAGAACTCGACCAATC 900  
QY 901 TCGGCTCTCAGGCTCGCATGGAAGGCTACAGGTCGCGCATCGAGGAAGTCTGTAAG 960  
Db 901 TCGGCTCTCAGGCTCGCATGGAAGGCTACAGGTCGCGCATCGAGGAAGTCTGTAAG 960  
QY 961 GATGTCGATATCTGTTATCATGACAGAACTCGGATATCATCTCTGTCATGATG 1020  
Db 961 GATGTCGATATCTGTTATCATGACAGAACTCGGATATCATCTCTGTCATGATG 1020  
QY 1021 GCCCAGATGAAGGATTAAGGCTATTGTCGTTAATCATCGGCCACTTCGATACGAATTCAT 1080  
Db 1021 GCCCAGATGAAGGATTAAGGCTATTGTCGTTAATCATCGGCCACTTCGATACGAATTCAT 1080  
QY 1081 ACAGATGGCTCATGAATACCCAGGCAATCAAGGACATCCCAATCAAGCCAGAAATACGAC 1140  
Db 1081 ACAGATGGCTCATGAATACCCAGGCAATCAAGGACATCCCAATCAAGCCAGAAATACGAC 1140  
QY 1141 ATGTGGGAATCCAGATGCGGACGATATCTCTTCTGCTGAGGCGCGCTCTTCTAAC 1200  
Db 1141 ATGTGGGAATCCAGATGCGGACGATATCTCTTCTGCTGAGGCGCGCTCTTCTAAC 1200  
QY 1201 CTTGGCTGCGCTACAGGTCAACCATCTTTGCTTATGTCATATGTCATTCACAAACAGACA 1260

Db 1201 CTTGGCTGCGCTACAGGTCAACCATCTCTTGGTATGTCATATGTCATTCACAAACAGACA 1260  
QY 1261 CTTGGCTGCGCTACAGGTCAACCATCTCTTGGTATGTCATATGTCATTCACAAACAGACA 1320  
Db 1261 CTTGGCTGCGCTACAGGTCAACCATCTCTTGGTATGTCATATGTCATTCACAAACAGACA 1320  
QY 1321 CCGAAGCATCTCGATGAAGAGTGCCTCGCTCCACCTCCAGTTCGATGTCGACCTT 1380  
Db 1321 CCGAAGCATCTCGATGAAGAGTGCCTCGCTCCACCTCCAGTTCGATGTCGACCTT 1380  
QY 1381 ACAAGCTTACACAGAGAGGCTGACTACATCAACGTTCCAGTTGAGGCTCTTTACAG 1440  
Db 1381 ACAAGCTTACACAGAGAGGCTGACTACATCAACGTTCCAGTTGAGGCTCTTTACAG 1440  
QY 1441 TCTGATGCTTACCGTTATTA 1461  
Db 1441 TCTGATGCTTACCGTTATTA 1461

RESULT 2  
US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 44.0%; Score 642.4; DB 14; Length 9025608;  
Best Local Similarity 66.0%; Pred. No. 1.5e-197;  
Matches 947; Conservative 0; Mismatches 481; Indels 6; Gaps 1;

QY 34 GAGTACAGAAATGCGGACATCAACCTCCATGTTCTCGGCGTGAAGAACTTACCCCTTCT 93  
Db 6119358 GACTTCAAGGTGCGCGGATCTCTCCCTGCGCGAGTTGCGGCGAAGGAGATCAACCTCGCC 6119477

QY 94 GAGAAGGAAATGCGGACATCAACCTCCATGTTCTTCGTAGCGTTATTCCGCTTCTAAGCCATTG 153  
Db 6119418 GAGCAGAGATGCGCGGCTGATGTCGATCCGCAAGGAGTACGCGGCGGAGGCGGCTC 6119477

QY 154 AAGGTGTCAAGATCTCTGTTCCCTCCATGATGACATGACATCCAGACAGCGCTCTCATGAG 213  
Db 6119478 GCGGCGCGCGCTCAACCGCTCCCTGCAATGACGCTGCGGCGAAGGAGTCACTCATGAG 6119537

QY 214 ACATCAACAGCTCTTGTGCTGATGTCAGATGGCTTCTTCCATCATCTTCTTACAA 273  
Db 6119538 ACCCTGCTGCGCTGCGGCGCGGAGTCCGCTTCTGCAACATCTTCTCCACCCAG 6119597

QY 274 GATACAGCGCTGCTGCTATGTCGTTGTCGCGCCCAACAGGCAACAGAGGCGCGCCGAG 333  
Db 6119598 GACCAGCGCGCGCGCTATCGGTCGGTCGCGGCAACGCGCGCGCGAGCCGAGGCG 6119657

Qy 334 ATCCAGTCTTCGCTCGAAGGCGGAACACTCCAGATACCTGGAGAGCAACATACGC 393  
Db 6119658 GTCCCGGCTTCGCTCGAAGGCGGAGACCTCGAGGAGTACTGGTGGTGCAGGAGAG 6119717  
  
Qy 394 GCTCTCACATGGCCAGATGTCAA-----GGCCCAACAGCAGAGTGTGCGATGATGGTGT 447  
Db 6119718 GCGCTGACTGGCGAACAACCCACCGGCGGCCGACATGATCTTGACGACGCGGT 6119777  
  
Qy 448 GATGTTACATCTCTCAATCTCAAGGCTTTCGAATTCGAAACAGCGGTCTTCCAGAG 507  
Db 6119778 GAGCGCAACCTCTCTGTCACAAAGGCGTTCGAGTACGAGAAGACGCAAGTTCGCTCC 6119837  
  
Qy 508 CCAACAGAAGCTGACACCTCGAATACCGTCTGCTTCTTCTACACTCAAGCAGTCTTC 567  
Db 6119838 GTCGACACCGCGGAGTCCGACGAGCACCGGCTCACTTCGAACTCTGACCGACCAATC 6119897  
  
Qy 568 AACCAAGACAAGAACCACTGGCACACAGTTGTCGCGGCATGAACGCTGTTTCCGAAGAG 627  
Db 6119898 ACCGACGCTCGCAAGATGGACCCAGCTCGGCTCGGAGATCCGGGCTGACCGAGAG 6119957  
  
Qy 628 ACAACAAAGTGTCCACCGCTCTACAGCTTCGAGAGAGGCGAAATCTCTCTTCCCA 687  
Db 6119958 ACCACGCGGCGTGCACCGCTGTACGAGATGCACCGGACGCGCACCTCTGTTCCG 6120017  
  
Qy 688 GCATCAACGTCACAGAGCTGTACAAAGTCCAAAGTTCGATTAACATCTACGGTCCGC 747  
Db 6120018 GGGATCAAGTGAACGAGCGCGTCAACAGTCCAAAGTTCGACCAAGTACAGGCTGCGC 6120077  
  
Qy 748 CACTCCCTTATCGATGTATCAACCGTCTCCGATGTCTATGATCGCGCGCAAGACAGCT 807  
Db 6120078 CACTCCCTGATCGAGCGCATCAACCGCGCCACCGAGCTCTGATCGCGCGCAAGACGCG 6120137  
  
Qy 808 CTCGTATGGTTAGCGGATGTCGCAAGGCTTCGCTCAATCTCTCGTGGGCCAAGGC 867  
Db 6120138 GTGCTCTCGGCTATGCGACGTGGGCAAGGCTGTGCGAGTCCCTGCGCGGCCAGGC 6120197  
  
Qy 868 GCTCGCTTATCAACAGAACCTGCAACCAATCTGCGCTCTCCAGGCTCCGATGAAGGC 927  
Db 6120198 GCGCCGCTGATGTCACCGAGATCGACCCGATCTGCGCCCTGCGAGCGCGATGAAGGC 6120257  
  
Qy 928 TACCAAGTCCCGCATCGAGGAAGTCTCAAGATGTCTGATATCTTCGTTACATGCACA 987  
Db 6120258 TACCAAGTCAAGCCCTCGACGAGTCTACGACAGGCGGACATCTTCTACACAGAG 6120317  
  
Qy 988 GGAATCTCGATATCTCTGTTGATGATGCGCCAGATGAAGATGAAGCTATGTC 1047  
Db 6120318 GGCACAAAGGACATCATATGCTCGGACATGGCCAGATGAAGCACCGAGCGATCTG 6120377  
  
Qy 1048 GGTAAATCGGCACTTCGATACGAATTTGATACAGATGGCTCATGAATACCGAGC 1107  
Db 6120378 GGCACATCGGCACTTCGACACGAGATCGACNTGGCGGCTTCGCGAGATCCCGGC 6120437  
  
Qy 1108 ATCAAGACATCCCAATCAAGCCAGAAATACGATGTGGAAATCCCGAGATGGCCAGCT 1167  
Db 6120438 ATGCTAAGGACGAGGTCAAGCGCAGGTCCACACTTGAAGTTCCTCCGCGCAAGTC 6120497  
  
Qy 1168 ATCTCTCTTCTGTTGAGGCGCGCTTCTTAACCTTGCTGCGTACAGTACCCCATCT 1227  
Db 6120498 CTCATGCTCTCTCCGAGGCGCGCTGTGAACCTTGGGCAACGCCACCGGCCACCGCTG 6120557  
  
Qy 1228 TTCTGTTATGTAATTCATTCACAAACAGACACTCTGCTCAGCTCGACCTCTACGAAG 1287  
Db 6120558 TTCTGTTATGTAATTCCTCTCGCGGACACAGACCTTGGCCAGATTCAGTGTTCACCAAG 6120617  
  
Qy 1288 AGAGGAATCTCGAAGAAAGTTTACACATTCGGAAGATCTCGATGAAGAATCGCT 1347  
Db 6120618 CCGGACGAGTACCGACCGAGCTTACGTGTGCGCCACAGACCTTCGACGAGAGTTCGCC 6120677  
  
Qy 1348 CGCTTCCACTCTGAGATCTTCGATGTCACCTTTACAAAGCTTACACAGAGAGGCTGAC 1407  
Db 6120678 CGCTTCCACTCTGAGTCTGCGGTGAAGTGAAGCTGACGAGCTTCGCGCCGAGCAGCGCG 6120737

Qy 1408 TACATCAAGCTTCAGTTGAGGTCCTTACAGTCTGATCGTTACCGTTATTAA 1461  
Db 6120738 TACATCGGCTCGAGGTCGAGGCGCCGTTCAAGTCGACCACTACCGCTACTGA 6120791  
  
RESULT 3  
US-10-156-761-5037  
; Sequence 5037, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 5037  
; LENGTH: 1455  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1455)  
US-10-156-761-5037  
  
Query Match 43.9%; Score 641.6; DB 14; Length 1455;  
Best Local Similarity 66.1%; Pred. No. 2e-199;  
Matches 945; Conservative 0; Mismatches 479; Indels 6; Gaps 1;  
  
Qy 34 GAGTACAGAAATGCGGACATCAACCTCCATGTTCTCGGCGGTAAAGAACTTACCTTGCT 93  
Db 25 GACTTCAAGGTGCGCGATCTCTCCCTGGCGGAGTTCGGCGCAAGGAGATCACTCCGCGC 84  
  
Qy 94 GAGAGGAAATGCGAGTCTTATGTTCTTCTGAGGCTTATTCGCTTCTAAGCATTG 153  
Db 85 GAGCAGAGATGCGCGGCTGATGTCGATCCGCAAGGAGTACGCGGAGCGGCGCCCTC 144  
  
Qy 154 AAGGCTGTCAGAATCTCTGGTCCCTCCACATGACAGTCCAGACAGCGCTCTCATCGAG 213  
Db 145 GCGGCGCGCGCTCACCGGCTCCCTGACATGACGCTGCGAGCCCGCTACTCATCGAG 204  
  
Qy 214 ACATCAAGCTCTTGGTGTGATGTGAGTGGGCTTCCTGCAACATCTTCTTACAA 273  
Db 205 ACCCTGTCGCGCTGGGCGCGAGGTTCGCTGGGCTCTCTGCAACATCTTCTCCACCGAG 264  
  
Qy 274 GATACAGCGCTGCTGATCTGTTGCGGCCCAACAGGCAACAGGAGGACGAGCGCT 333  
Db 265 GACACGCGCGCGCGCTATGCGGTTCGGCGCGCAACGCGCACCGCCGAGCGCGC 324  
  
Qy 334 ATCCAGCTCTTTCGCTGGAAGGCGGAAACACTCCAGAACTACTGGGAGAAACACATACGCG 393  
Db 325 GTCCCGGCTCTTCGCTGGAGGCGGAGCCCTGGAGGAGTACTGGTGGTGCACGAGAG 384  
  
Qy 394 GTCTCACATGGCGAGATGGTCA-----AGGCCCAACAGCAGGTTGTGATGATGGTGGT 447  
Db 385 GCGCTGACCTGGCGGAACACCCCAACCGGCGCGGCAACATGATCTCTGACGAGCGGCT 444  
  
Qy 448 GATGTTACATCTCTCATCTTCGAGGCTTCGAATTCGAACAGCGGCTGCTGTTCCAGAG 507  
Db 445 GAGCGCAACCTCTCTGCTCCACAGGCGCTCGAGTACGAGAAGGAGCGCAAGGTTCCGCTCC 504  
  
Qy 508 CCAACAGAAGCTGACACCTCGAATACCGTTCGCTTCTTCTACTCAAGCAGGTCTTC 567

Db 505 GTGCACACGGCGAGTCGAGGACCGCGTCATCTCTGACCTCTGCAACGACCATC 564  
Qy 568 AACCAAGACAGACCACTCGGCAACAGTTGTCGGGCGATGAAACGGTGTTCGGAAG 627  
Db 565 ACCGACGGCTCGGAGAGTGAGCCCGAGCTCGCTCGGAGATCGCGCGCTGACCGAGG 624  
Qy 628 ACAACAAAGAGTGTCCACCGCTCTACAGCTCGAGAGGAGGCGAACTCTCTTCCCA 687  
Db 625 ACACAGCGGGTGACCCCTGTGACGATGACACCGGACCGCTCTCTGTTCGG 684  
Qy 688 GCCATCAACGTCACGACGCTGTTCACAAAGTCCAAAGTTCGATAACATCTACGGCTCGCG 747  
Db 685 GCGATCAAGTGAACGACGCGCTACCAAGTCTGAAGTTGACACAAAGTACGGCTCGCG 744  
Qy 748 CACTCCCTTATCGATGTTATCAACCTGTCTCCGATGTCATGATCGGCGGCGACGCT 807  
Db 745 CACTCCCTGATCGACGCGCATCAACCGCGCACCGAGCTCTCTGATCGGCGGCGAAGCCGG 804  
Qy 808 CTCGTATGGTGTACGCGATGTGCGCAAGGCTCGGCTCAATCCCTCGTGGCGCAAGGC 867  
Db 805 GTGCTCTCGGCTATGCGACGTGGGCAAGGCTGTGGGAGTCCCTGCGCGGCCAGGGC 864  
Qy 868 GTCGCTTATCATCAGCAACATCGACCAATCTCGCTCTCCAGGCTGCGCATGGAAGGC 927  
Db 865 GCGCGGTGATCGTCACGAGATCGACCGATCTGCGCCCTGCGGCGGCGATGACGCGC 924  
Qy 928 TACGAGTCCGCGCATCGAGGAGTCTGCAAGGATGCTATCTCTGTTATCATGACAC 987  
Db 925 TACGAGTCAAGCCCTCGACGAGGTCATCGACAGGCGGACATCTCTGTCACCGACG 984  
Qy 988 GGAATCTCGATATCATCTCTGTGATGATGATGCGGCGGATGAGGATAGGCTATTC 1047  
Db 985 GGCACACAGGACATCATCTGCTCGGACATGCGGCAAGATGAAACACAGGCGATCGTG 1044  
Qy 1048 GGTAACTCGGCGCATCTCGATACGAAATTCAGATGAGTGGCTCATGAAATACCGAGC 1107  
Db 1045 GGCACATCGGCGCATCTCGACACAGGATCGACATGCGCGGCTTGGCGAGATCCGCGC 1104  
Qy 1108 ATCAAGCACATCCCAATCAAGCAGAAATAGACATGTTGGAAATTCGAGATGCGCACGCT 1167  
Db 1105 ATCGTCAAGGACGAGGTCAGGCGGAGTCCACATCTGGAAGTTCCCGACGCGAAGTC 1164  
Qy 1168 ATCTCTCTTGTGAGGCGCGCTTCTTAACCTTGGCTGCGTACAGTCAAGTCAACCATCT 1227  
Db 1165 CTCACTGCTCTCTCGAGGCGCGCTCTGTAACCTTGGGCAAGCCACCGCGCACCGCTCG 1224  
Qy 1228 TTGTTATGTAATGTCATTCACAAACAGACATCGCTCAGCTCGACCTCTACGAAAG 1287  
Db 1225 TTGTTGATGTCCAACTCTTTCGCGGACAGACCTTGGCCAGATCGAGCTGTTTACCAAG 1284  
Qy 1288 AGAGGAATCTCGAGAGAGGTTTACACATTCGAGAGCATCTCGATGAAGAATCGCT 1347  
Db 1285 CCGACAGATACCGGACCGAGCTTACGTGCTGCGGCAAGCACTTCGACGAGAGTGCGCC 1344  
Qy 1348 CGCTCCACCTCGGATCTCTCGATGTCCACCTTACAAAGTTTACACAGAGCAGCTGAC 1407  
Db 1345 CGCTCCACCTCGGATCTCTCGGCTGAGAGCTGACAGCTTCCGCCCGGAGGAGCGCG 1404  
Qy 1408 TACATCAAGTTCAGGAGGTTCTTACAAAGTCTGATGTTTACCGTTA 1457  
Db 1405 TACATCGGCTGAGGTTCGAGGCGCGTTCAGTTCGAGACCTACCGCTA 1454

## RESULT 4

US-09-746-660A-97  
; Sequence 97, Application US/0974660A  
; Publication No. US20030049804A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompeius, Markus  
; APPLICANT: Kroger, Burkhard  
; APPLICANT: Schroder, Hartwig  
; APPLICANT: Zelder, Oskar  
; APPLICANT: Habershauer, Gregor

; APPLICANT: Kim, Jun-Won  
; APPLICANT: Lee, Heung-Schick  
; APPLICANT: Hwang, Byung-Joon  
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING  
; TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS  
; FILE REFERENCE: BGI-121CP2  
; CURRENT APPLICATION NUMBER: US/09/746,660A  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 09/606740  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 09/603124  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 60/141031  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: 60/142101  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: 60/148613  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/187970  
; PRIOR FILING DATE: 2000-03-09  
; PRIOR APPLICATION NUMBER: DE 19931420.9  
; PRIOR FILING DATE: 1999-07-08  
; NUMBER OF SEQ ID NOS: 125  
; SOFTWARE: PatentIn Vers. 2.0  
; SEQ ID NO 97  
; LENGTH: 1557  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (101)..(1534)  
; OTHER INFORMATION: RXN00132  
US-09-746-660A-97

Query Match 40.8%; Score 596.4; DB 11; Length 1557;

Best Local Similarity 64.5%; Pred. No. 1.4e-184;

Matches 923; Conservative 0; Mismatches 501; Indels 6; Gaps 2;

Qy 32 TGAAGTACAGAAATTCGCGACATCAACCTCCATGTTCTCGGCGGTAAAGAACTTACCCCTTG 91  
Db 114 TGAAGTTCAGAGTTGCGGATCTTTCATAGCAGAGGAGGACCTCACCAGATTTCGCTTG 173  
Qy 92 CTGAGAGGAATCCAGGCTTATGTTCTTCGTGAGCGTTATTCGGCTTCTAAGCCAT 151  
Db 174 CAGATGATGAGATGCCAGGCTTCATGCAAGTTCGCGCAAGAAATTCGACAGAGGAGCCCTT 233  
Qy 152 TGAAGGGTGTGAGAAATCTCTGGTTCCCTCCACATGACAGTCCAGACAGCGCTCCTCATCG 211  
Db 234 TGAAGGGCGCCGAAATTCGCTGTTCTATCCACATGACGTTCCAGACCGCGCTTATTG 293  
Qy 212 AGACACTCAGCTCTTGGTGTGATGTCAGATGGGTTCTCTGCAACATCTTCTCTACAC 271  
Db 294 AGACCTCCTCCTGCTTGGGCGCTGAGTTCTGTTGGGCTCTCTGCAACATTTTCTCCACC 353  
Qy 272 AAGATACAGCGCTGCTGCTATCTGTTGTCGCCCCAAGCAGCACACAGAGAGAGCCAGCG 331  
Db 354 AGATGAGGCTGCGAGGCTATCGTTGTCGGC---TCCGGCACCGTTCGAGAGGAGCGTG 410  
Qy 332 GTATCCCACTCTTCGCTTCGAGGCGGAAACACTCCAGAAATCTGGGAGAACATACACC 391  
Db 411 GTGTTCCAGTATTCGCTGGAAGGTTGAGTCACTGGAGGAGTACTGCTGTCATCAACC 470  
Qy 392 GCGCTCTCAGATGCGCAGATGTTCAAGGCCACAGAGGTTGTCGATGATGTTGTTGATG 451  
Db 471 AGATCTTCAGTGGGCGATGAGC---TGCCAAACATCATCTCCGACGAGCGGCTGAGC 527  
Qy 452 CTACACTCTCATCTCAAGGCTTCGAATTCGAAACAGCGCGTCTGTTCCAGAGCCAA 511  
Db 528 CCACCATCGCTGTTATTCGCGGTTCGGAATACGAGGAGGCTGTTCTGTTCCACAGCAG 587  
Qy 512 CAGAAGCTGACAACTCGAATACCGCTGCTTCTTGTCTACACTCAAGCAGGTTCTCAACC 571  
Db 588 AGGCCAACGATTCGATGAGTCAATTCGTCATCTTGGGCAATGCTCGCTGAGGTTCTGCTG 647

QY 572 AAGACAGAACCACTGSCACACAGTTGCTGCGGATGAGACGTTGTTCCGAGAGACAA 631  
Db 648 CAGAGCCCTGGCAAGTGGGCAAGATCGCTGAGGCGGTAAAGGGTGTCAACGAGGAAACA 707  
QY 632 CAACAGGTGTCCACCGCTCTACACAGCTCGAGAGAGGCAAACTCTCTCCAGACCA 691  
Db 708 CCACCGGTGTGACACCGCTCTACACAGCTCGAGAGAGGCAAACTCTCTCCAGGGA 767  
QY 692 TCAAGCTCAACAGAGCTGTGTACAAAGTCCAAAGTTCGATTAACATCAAGCTCCCGCCACT 751  
Db 768 TGAAGCTCAACAGAGCTGTGTACAAAGTTCGATTAACATCAAGCTCCCGCCACT 827  
QY 752 CCTTATCGATGATCAACCGTCTTCGATGTCATGATCGGCGCAAGACAGCTCTCG 811  
Db 828 CCTGATCGAGGATCAACCGGCGCACTGACATGCTCATGGGCGCAAGACAGCTCTG 887  
QY 812 TCATGGGTTCAGGCGATGTGCGCAAGGCTGGCTCAATCCCTCCGTCGCAAGCGCTC 871  
Db 888 TCTGCGGTTCAGGCGATGTGCGCAAGGCTGGCTCAATCCCTCCGTCGCAAGCGCTC 947  
QY 872 GCGTTATCATACAGACTCGACCCAAATCTGCGTCTCCAGGCTTCGATGAGAGGCTACC 931  
Db 948 GCGTCAAGGTCAACGAGCTGACCCAAATCAAGCTCTTCAGGCTCTGATGAGGCTACT 1007  
QY 932 AGTCCGCGCATCGAGAGTCTCAAGGATGTCGATCTTCGTTATCATGCAAGGAA 991  
Db 1008 CTGTGTCACCGTTCAGGAGGATCGAGGAGCGGACATCGTGATCACCGGACCGCA 1067  
QY 992 ACTGGGATCATCTCTGTTGACATGATGCGCCAGATGAAGATTAAGGCTATTGCGTA 1051  
Db 1068 ACAAGGACATCATCTCTCGACAGATGCTCAAGATGAGGATCAAGCTCTGCTGGCA 1127  
QY 1052 ACATCGGCACTTCGATTAACGAAATGATGATGAGGCTTCGATGAGGATCAAGGCAATCA 1111  
Db 1128 ACATCGGCTCACTTTGATTAATGATGATGATGATGATGATGATGATGATGATGAT 1187  
QY 1112 AGCAGATCCCAATCAAGCAGAGATGACATGTCGGAATCCAGATGCGCAGCTATCC 1171  
Db 1188 CCGCAGACAGATCAAGCAGAGTTCAGGATGACAGGTTCACTCTCAAGGTCGCTCATCA 1247  
QY 1172 TCCTTTTTCGAGGCGGCTCTTTTAACTTTGGCTGCTGATGAGGTCACCATCTTTTCG 1231  
Db 1248 TGCTCTGTCCGAGGTCGCTCTTGAACCTTGGCAGCCACCGAGACCATCATTTG 1307  
QY 1232 TTATGTCATGTCATCAAAACAGACACTCGCTCAGCTGACCTCTACGAAAGAGAG 1291  
Db 1308 TCATGTCCAACTCTTTGCGCGATCAGACCATTTGCGCAGATCGAACTGTTCCAAAACGAAG 1367  
QY 1292 GAAATCTCGAGAGAGGTTTACACACTTCGGAAGCATCTCGATGAAGAGTCGCTCGCC 1351  
Db 1368 GACAGTACGAGACGAGTCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1427  
QY 1352 TCACTCGGATCTCTGATGTCCACTTTACAAAGCTTTACAGAGAGGAGGCTGACTACA 1411  
Db 1428 TCAAGTGTGAGGCTCTCGGCGTCACTCAACGAACTGACAAAGGAGGAGGCTGAGTACA 1487  
QY 1412 TCAAGTGTGAGGCTCTCGGCGTCACTCAACGAACTGACAAAGGAGGAGGCTGAGTACA 1461  
Db 1488 TGGCGTTGAGGTTGAGGCGGCAATTCAGCGGAGGAGGCTACCGGCTACTAA 1537

RESULT 5

US-09-919-854-1  
; Sequence 1, Application US/09919854  
; Publication No. US20030100080A1  
; GENERAL INFORMATION:  
; APPLICANT: FARWICK, MIKE  
; APPLICANT: HUTHMACHER, KLAUS  
; APPLICANT: BREHME, JENNIFER  
; APPLICANT: PFERLE, WALTER  
; APPLICANT: BINDER, MICHAEL  
; APPLICANT: GREISSINGER, DIETER

; APPLICANT: THIERBACH, GEORG  
; TITLE OF INVENTION: Nucleotide sequences which code for the sahH gene  
; FILE REFERENCE: 211739US0X  
; CURRENT APPLICATION NUMBER: US/09/919,854  
; PRIOR FILING DATE: 2002-12-30  
; PRIOR APPLICATION NUMBER: DE 100 44 706.6  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: DE 101 09 685.2  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 60/294,277  
; PRIOR FILING DATE: 2001-05-31  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 1939  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (227)..(1720)  
; OTHER INFORMATION:  
; US-09-919-854-1

Query Match 40.8%; Score 596.4; DB 11; Length 1939;  
Best Local Similarity 64.5%; Pred No. 1.6e-184;  
Matches 923; Conservative 0; Mismatches 501; Indels 6; Gaps 2;

QY 32 TCGAGTACAGAAATGCGACATCAACCTCCATGTTCTCGCCGTAAGAACTTACCCCTG 91  
Db 300 TGGACTTCAAGTTGCGGATCTTCACTAGCAGAGGAGGAGCTACACAGATTCTGCTTG 359  
QY 92 CTGAGAGGAAATGCCAGGCTTATGTTCTTCTGAGCGTTATTCGCTTCTAAGCAT 151  
Db 360 CAGATATGATGATGCCAGGCTTCATGCAATGCGCAAGAAATTCGAGACGAGCGCTT 419  
QY 152 TGAAGGTTGTCAGAAATCTCTGTTCCCTCCACATGACATGACAGCAGCGCTCTCATCG 211  
Db 420 TGAAGGCGCCGAAATGCTGTTCTATCCACATGAGGTCAGACCGCGCTTATTG 479  
QY 212 AGACATCAGAGCTTCTGTTGCTGATGTCAGATGGGTTCTCTGCAACATCTTCTACAC 271  
Db 480 AGACCTCACTGTTTGGCGCTGAGGTTCTGTTGGGTTCTCTGCAACATTTTCTCACCC 539  
QY 272 AGATACAGCGCTGCTGTTATGTTGCTGTCGSCCAACAGGCGACACAGAGAGCGCG 331  
Db 540 AGATAGGCTGCGAGCGGCTATCGTTGTCGCG---TCGCGCACCTTCGAGAGCGAGCTG 596  
QY 332 GTATCCAGTCTTTCGCTGGAAGGCGGAAACACTCCAGAAATACTGGGAGAACACATACC 391  
Db 597 GTGTTCCAGTATTCGCTGGAAGGCTGAGTCACTGGAGGAGTACTGTTGTCATCAACC 656  
QY 392 GCGCTCTCAGATGCGAGATGTCAGAGCCCAACAGAGGTTGTCGATGATGTTGTTGATG 451  
Db 657 AGATCTTCAGTGGGCGATGAGC---TGCCAAACATGATCTCTCAGACGCGGCTGACG 713  
QY 452 CTACACTCTCATCTCAAGGCTTCGAAATTCGAAACAGCGGCTGCTCTCCAGAGCAA 511  
Db 714 CCACATGCGCTGTTATTCGCGGTCGGAATACGAGCGGCTGTTGTTTCCACAGAG 773  
QY 512 CAGAAAGTGAACAACCTCGAATACCGCTGCTTCTTGTCTACACTCAAGCAGGCTTCTCAACC 571  
Db 774 AGGCCACGATTCGATGATGATACATCGCATTTCTTGGGATGCTGCGTGAAGTTCTTCTG 833  
QY 572 AAGACAGAAACCACTGGCAACAGTTGTCGCCGATGAAACGGTGTTCGGAAGAGCAA 631  
Db 834 CAGAGCTTGGCAAGTGGGCAAGATCGCTGAGGCGGTTAAGGCTGTCCACCGAGAAACCA 893  
QY 632 CAGAGGTTGCCCGCTTACAGCTCCGAGAGGAGGCGAACTCTCTTCCAGCCA 691  
Db 894 CCACCGGTGTCACCGCTGTACCACTTCGCTGAAGAGGCGTGTGCTTCCAGCGA 953  
QY 692 TCACGTCACAGCGCTGTTTACAAAGTCCAAAGTTCGATTAACATCTACGCTGCGGCCACT 751

Db 954 TGAAGCTCAACGACGCTGTCCACCAAGTCCAGTTTGATACAAAGTACGGACCCGCCACT 1013  
Qy 752 CCTTATCGATGATCAACCGTGTCTCCGATGTCATGATCGGCGGCAAGACAGCTCTCG 811  
Db 1014 CCTGATCGACGGCATCAACCGCGCTCATGCTCATGCGGCGCAAGAACGCTGCTTG 1073  
Qy 812 TCATGGTTACGGCGATGTCGGACAGGCTGGCTCAATCCCTCCGTGGCCAGGCGCTC 871  
Db 1074 TCTGGGTTACGGCGATGTCGGCAAGGCTGGCTCAATCCCTCCGTGGCCAGGCGCTC 1133  
Qy 872 GGGTTATCATCAAGAACTCCACCAATCTGGCTCTCCAGGCTCCAGGCTGCGATGGAAGCTACC 931  
Db 1134 GGGTCAAGGTACCGAAGCTGACCAATCAAGCTCTTCAGGCTCTGATGGATGCTACT 1193  
Qy 932 AGGTCCGCGCATCGAAGAACTGTCGATGATGTCGATATCTTCGTTATCATCAAGAA 991  
Db 1194 CTGTGTCACCGTTGATGAGCCATCGAGGACCGCGCATCGTGATCACCGGACCGCA 1253  
Qy 992 ACTGGATATCATCTCTGTTGACATGATGGCCCGACATCAAGGATAGGCTATGTCGTA 1051  
Db 1254 AAGAGCATCATTTCTTCGACGATGCTCAAGATGAAGGATCAAGCTCTGTCGGCA 1313  
Qy 1052 ACATCGGCACTTCGATAAGAAATGATACAGATGGCTCATGAAATACCCAGGATCA 1111  
Db 1314 ACATCGTCACTTTGATAATGATCGATGATGCAATCCCTGTTGACCGCGACGATCA 1373  
Qy 1112 AGCATCCGATCAAGCCAGAAATACGATGTCGGAATTCGACATGTCGACGATGTCGATGCC 1171  
Db 1374 CCGCACACGATCAAGCCACAGTCCGAGATTCACCTTCGACCGGCTGCTCCATCA 1433  
Qy 1172 TCTTCTTGTGAGGCGGCTCTTTAACTTTGGCTGGCTACAGTCAAGCTCAAGCTTTTCG 1231  
Db 1434 TGGTCTGTCGAAGTCTGCTTTGAACCTTTGGACGACCGACACCCATCTTG 1493  
Qy 1232 TTATGTCATGTCATCAAAACACACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1291  
Db 1494 TCATGTCACCTTTTCGCGATCAGACCAITTCGCGATCGAATGTTCCAAACGGAAG 1553  
Qy 1292 GAAATCTCGAAGAAAGTTTACACACTTCGAGACATCTCGATGAAGAGTCTGCTGCC 1351  
Db 1554 GACATGACGAGACGAGGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1613  
Qy 1352 TCCACTCGGATCTCTGATGCTCCACTTACAAAGCTTACACAGAGGAGGCTGACTACA 1411  
Db 1614 TCCAGTTGAGCTCTCTCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1673  
Qy 1412 TCAAGTTCCAGTTGAGGCTCTTACAGTCTGATGCTTACCGTTATTAA 1461  
Db 1674 TCGGGTTGAGGTTGCGAGGCCATTCAGCGGAGCACTACCGCTACTAA 1723

RESULT 6  
US-09-738-626-1  
; Sequence 1, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATBISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738, 626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162

; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO. 1  
; LENGTH: 3309400  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-1  
  
Query Match 40.8%; Score 596.4; DB 10; Length 3309400;  
Best Local Similarity 64.8%; Pred. No. 1e-182;  
Matches 923; Conservative 0; Mismatches 501; Indels 6; Gaps 2;  
  
Qy 32 TCGAGTACAGAAATGGCGACATCAACCTCCATGTTCTCGGCGCTAAGGAACTTACCCCTTG 91  
Db 788673 TGGACTTCAAGTTGCGATCTTCTACTAGCAGGCGAGGAGCTCACCAGATTCGTCTTG 788732  
Qy 92 CTGAGAAGAAATGCCAGGTCTTATGTTCTTCTGAGCGTTATTCGCTTCTTAAAGCAAT 151  
Db 788733 CAGAGTATGATGCCAGGTCTCATGCAATTGCGCAAGGAAATTCGACAGCAGAGCCTT 788792  
Qy 152 TGAAGGTTGTCAGAAATCTCTGTTCCCTCCACATGACAGTCAGACGCGCTCTCATCG 211  
Db 788793 TGAAGGCGCCGAAATGCTGTTCTTATCCATGAGGTCAGACCGCGCTGCTATTG 788852  
Qy 212 AGACATCTACAGCTCTTGGTGTGATGTCAGATGGGCTTCTGCAACATCTTCTTACAC 271  
Db 788853 AGACCTCACTGTTTGGCGCTGAGGTTCTGTTGGGCTTCTGCAACATTTCTCCACCC 788912  
Qy 272 AAGATACAGCGCTGCTGCTGATCTGTTCTGCGCCCAAGAGGACACACAGAGAGCAGCG 331  
Db 788913 AGATGAGGCTGCGAGCGCTATCTGTTGTCGCGC--TCGCGCACCGCTCGAAGAGCGAGCTG 788969  
Qy 332 GTATCCCAAGTCTTCGCTCGAAGGCGGAAACACTCCAGAACTACTGGGAGAACACATACC 391  
Db 788970 GTGTTCCAGTATTCGCGTGGAGGCTGATCTCTGGAGGACTACTGTTGCTGATCAAC 789029  
Qy 392 GGGCTCTACATGCCAGATGTCGAGGCCACAGAGAGGTTGTCGATGATGTTGGTGAATG 451  
Db 789030 AGATCTTCAGCTGGGCGATGAGC--TGCCAAACATGATCTCCAGACAGCGCGTGAAG 789086  
Qy 452 CTACACTCTCATCTCCAGGCTTCGAAATTCGAAACAGCGGCTGCTGTTCCAGAGCAA 511  
Db 789087 CCACCATGGCTGTTATTCGCGTTCGAAATACGAGAGGCTGCTGTTCCACAGCAG 789146  
Qy 512 CAGAAGCTGACAACTCGAATACCGCTGCTGTTCTTGTACACTCAAGCAGGCTCTTCAACC 571  
Db 789147 AGGCCAACGATTCGGATGATGATCATCGCATTTCTGGGCAATGCTGCGTGAAGTTCTTGCTG 789206  
Qy 572 AAGACAGAACCACTGCGACACAGTGTGCTGCGGCATGAACGGTGTTCGGAAGAGCAA 631  
Db 789207 CAGAGCTTGGCAAGTGGGCAAGATCGCTGAGGCGGTTAAGGCTGTCCACGAGAAACCA 789266  
Qy 632 CAACAGGTGTCACCGCTCTACAGCTCGAGAGGAGGAGGCAAACTCTCTTCCAGCA 691  
Db 789267 CCACCGGTGTCACCGCTGTACCACTTTCGCTGAAGAGGCTGCTGCTTTCACAGCA 789326  
Qy 692 TCACAGTCAACGAGCTGTGTACAAAGTCCAAAGTTCGATTAACATCTACGCTGCGGCCACT 751  
Db 789327 TGAAGTCAACGAGCTGTGTACCAAGTCCAAAGTTCGATTAACAGTACGCGACCGCCACT 789386  
Qy 752 CCCTATCGATGATCAACCGCTGCTCCGATGTCGAGTATGATGTCGGCGGCAAGACGCTCG 811  
Db 789387 CCCTGATCGAGCGCATCAACCGCGCTCATGACATGCTCATGGGCGGCAAGACGCTGTTG 789446  
Qy 812 TCATGGTTACGGGATGTCGCGAGGCTGCGCTCAATCCCTCCGTCGCGCAAGCGCTC 871  
Db 789447 TCTCGGTTACGGGATGTCGCGAGGCTGCGCTGAGGCTTTCGACGCGCAGGCGCTC 789506  
Qy 872 GCGTTATCATCACAGAACTCGACCCCAATCTGCGCTCTCCAGGCTGCGCATGGAAGGCTACC 931

Db 789507 GCGTCAAGGTCAACGAAGCTGACCCAAATCAACGCTCTTCAGGCTCTGTAGTGGCTACT 789566  
Qy 932 AGGTCCGCGCATCGAAGAAAGTGTCAAGGATGTGATATCTTCGTTCATGCAACAGGAA 991  
Db 789567 CTGTGGTCAACGTTGATAGGCCATCGAGGACGCCGACATCGTGATCACCGGACCGGCA 789626  
Qy 992 ACTCGGATATCATCTCTGTGATGATGATGCGGCCAGATGAAGTAAAGTATGTGCGTA 1051  
Db 789627 ACAAGGATCATCTTCCTTCGACAGATGCTCAAGATGAAGGATCACGCTCTGCTGGGCA 789686  
Qy 1052 ACATCGGCACTTCGATAACGAATTTGATACAGATGGCTCATGAATATCCAGGCATCA 1111  
Db 789687 ACATCGGTCACTTTGATAATGATCGATATGCAATTCCTGTTCACCGCGACGATCA 789746  
Qy 1112 ACACATCCAAATCAAGCCAGATACGATGCGGGAATTCACAGATGGCCACGCTATCC 1171  
Db 789747 CCGGACACAGCATCAAGCCACAGGTGCGAGGTTCACTTCTCCACCGTCTGCTCCATCA 789806  
Qy 1172 TCTTCTTGTGAGGCGCGCTTCTTAACCTTTGGCTGGCTACAGTCAACCATCTTTCG 1231  
Db 789807 TGTCTCTTCGAAGTTCGCTTTGAACCTTTGGCAAGCCACCGACACCCCTCATTTG 789866  
Qy 1232 TTATGCTAATGTCAATCAAAACGACACTCGCTTCAGCTCGACCTTACGAAAGAGAG 1291  
Db 789867 TCATGCTCAACTCTTTGCGGATCAGACCATTTGGCAGATCGAATCTTCCAAACGAAG 789926  
Qy 1292 GAAATCTCGAAGAGGTTTACACACTTCGAGAGCTCTGATGAAAGTTCGCTCGCC 1351  
Db 789927 GACAGTACGAGACGAGGTTACCGTCTGCTTAAGGTTCTCGACGAAAGGTTGGCAGCA 789986  
Qy 1352 TCCACTCGGATCTCTGATGTCACCTTTACAAAGCTTTACAGAGAGGCTGACTACA 1411  
Db 789987 TCCAGTTGAGGCTCTCGGCGGTGACCTCACGAACTTACAGAGAGGCTGATGACA 790046  
Qy 1412 TCAAGTTCAGTTCAGGTCCTTCAAGTCTGATGCTTACCGTTATTAA 1461  
Db 790047 TCGGGGTTGACGTTGAGGCGCCATTCAAGCGGAGCATACCGCTACTAA 790096

## RESULT 7

US-09-738-626-836  
; Sequence 836, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738, 626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 836  
; LENGTH: 1422  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-836

Query Match 40.7%; Score 594; DB 10; Length 1422;

Best Local Similarity 64.5%; Pred. No. 8.3e-184;  
Matches 920; Conservative 0; Mismatches 500; Indels 6; Gaps 2;  
Qy 32 TCGAGTACAGAAATTCGCGACATCAACCTCCATGTTCTCGGCCGTGAAGAACTTACCTTTG 91  
Db 2 TGGACTTCAAGTTGCGATCTTTCACTAGCAGGCGAGACGTACACAGATTGCTTTG 61  
Qy 92 CTGAGAGGAAATTCGCGAGTCTTATGTTCTTCTGAGGCTTATTCGCTTCTTAAGCCAT 151  
Db 62 CAGATATGAGATGCCAGTCTCATGCAATTCGCAAGGAATTCGCGACAGCAGCGCTT 121  
Qy 152 TGAAGGTGTGCAAAATCTGTTCTCCATGACAGTCAAGTCAAGAGCCGCTCTCATCG 211  
Db 122 TGAAGGCGGCGCAATTCGTTCTTATCCATGACGTTCCAGACCCGCGTGTATTG 181  
Qy 212 AGACACTCACAGTCTTGTGCTGATGTCAGATGGGCTTCTGCAACATCTTCTCTACAC 271  
Db 182 AGACCTCACTGCTTTGGGCGCTGAGGTTCTTGGGCTTCTGCAACATTTCTTCCACC 241  
Qy 272 AAGATACAGCGCTGCTGCTATCTGTTGTGCGCCCAACAGSCACACAGAGAGCCAGCG 331  
Db 242 AGGATGAGGCTGCGAGCGGCTATCGTTGTGCGC---TCCGCGCACCGTTCGAAGAGCCAGCTG 298  
Qy 332 GTATCCCAAGTCTTCGCTGGAAGGCGAAACACTCCAGAACTACTGGGAGAACACATACC 391  
Db 299 GHTTCCAGTATTCGCGTGGAGGCTGAGTCACTGGHAGGACTGTTGGTGTGCTACACC 358  
Qy 392 GCGCTCTCAATGCCAGATGCTCAAGGCCACAGCAGAGTGTTCGATGATGTTGTTGATG 451  
Db 359 AGATCTTCAGCTGGGCGATGAGCT---GCCAACAATGATCTCTCGACGACGCGGTGACG 415  
Qy 452 CTACACTCTCATCTCGAAGGCTTCGAATTCGAACAGCGGCTGCTTTCAGAGCCAA 511  
Db 416 CCACCATGCGTGTATTTCGCGTTCGCGAATAAGAGAGGCTGCTGTTGTTCCACAGCAG 475  
Qy 512 CAGAAGTGTCAACCTCGAATACCGCTCGTTCTTGTCTACACTCAAGCAGAGTCTTCAACC 571  
Db 476 AGGCCAAGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 535  
Qy 572 AAGCAAGAACCACTGCGACACAGTGTGTCGCGCATGAACGTTTCCGAGAGACAA 631  
Db 536 CAGAGCTTGGCAAGTGGGCAAGATCGCTGAGGCGGTTAAGGCTGTCCACGAGGAAACCA 595  
Qy 632 CAAAGGTGTCCACCGCTCTACAGCTCGAGAGAGGCGGCAAACTCTCTTCCAGCCA 691  
Db 596 CCACCGGTGTCCACCGCTGTACCACTTCGCTGAAGAGGCGTGTGCTTCTCCACGCA 655  
Qy 692 TCAAGCTCAACGACGCTGTTACAAAGTCCAGTTCGATTAACATCTACGCTGCGCCACT 751  
Db 656 TGAAGCTCAACGAGCTGTTCACAAAGTCCAGTTCGATTAACAGTACGCGACCGCCACT 715  
Qy 752 CCCTTATGATGATCAACCGTGTTCGATGTCATGATGTCGCGCGGCAAGACAGCTCTCG 811  
Db 716 CCCTGATCGAGCGCATCAACCGCGCTACGATGCTCATGCGCGCAAGAACGCTGTTG 775  
Qy 812 TCATGGGTGAGGCGATGTCGAGAGGCTGCGCTCAATCCCTCGTGGCGCAAGCGCTC 871  
Db 776 TCTCGGTACCGCGATGTCGCGAAGGCTGCGCTGAGGCTTTTCGACGCGCGAGGCGCTC 835  
Qy 872 GCGTTATCATCACAGAACTCGACCAATCTGCGCTCTCCAGGCTGCCATGGAAGGCTACC 931  
Db 836 GCGTCAAGGTCAACGAGCTGACCAATCAACGCTCTTCAGGCTCTGATGATGCTACT 895  
Qy 932 AGGTCCCGCCATCGAGGAGTGTCAAGGATGTGATATCTTCTGTTACATGCAAGAA 991  
Db 896 CTGTGCTCACCGTTGATGAGGCCATCGAGGACGCGACATCGTGATCACCGGACCGCA 955  
Qy 992 ACTGGATATCATCTCTGTTGATGATGATGCGCCAGATGAGGATGAGGCTATTTCGCTA 1051  
Db 956 ACAAGGACATCATTTCTTCGAGCAGATGCTCTAGATGAAGGATCACGCTCTGCTGGCA 1015  
Qy 1052 ACATCGGCCACTTCGATAACGAATTTGATACAGATGCGCTCATGAAATACCCAGGCATCA 1111





Db 130425 CACCTCGAGCCCTCGCGCTGAAGCTGACGACGCTGCGCCCGGAGCAGCCTCTGTACATC 130366  
QY 1414 AACGTTCCAGTTGAGGCTCTTACAAAGTCTGATGCTTACCGTTATTAA 1461  
Db 130365 GCGCTAGAGTCTGACGCCCGCTACAAGCCGGACCACTACCGCTACTGA 130318

RESULT 9  
US-09-746-660A-101  
; Sequence 101, Application US/09746660A  
; Publication No. US20030049804A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompejus, Markus  
; APPLICANT: Krogger, Burkhard  
; APPLICANT: Schröder, Hartwig  
; APPLICANT: Zelder, Oskar  
; APPLICANT: Haberhauer, Gregor  
; APPLICANT: Kim, Jun-Won  
; APPLICANT: Lee, Heung-Schick  
; APPLICANT: Hwang, Byung-Joon  
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING  
; TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS  
; FILE REFERENCE: BGI-121CP2  
; CURRENT APPLICATION NUMBER: US/09746, 660A  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 09/606740  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 09/603124  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 60/141031  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: 60/142101  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: 60/148613  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/187970  
; PRIOR FILING DATE: 2000-03-09  
; PRIOR APPLICATION NUMBER: DE 19931420.9  
; PRIOR FILING DATE: 1999-07-08  
; NUMBER OF SEQ ID NOS: 125  
; SOFTWARE: Patent In Vers. 2.0  
; SEQ ID NO 101  
; LENGTH: 1396  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (101)..(1396)  
; OTHER INFORMATION: FRXA01371  
US-09-746-660A-101

Query March 36.2%; Score 529.6; DB 11; Length 1396;  
Best Local Similarity 64.3%; Pred. No. 1.1e-162;  
Matches 828; Conservative 0; Mismatches 454; Indels 6; Gaps 2;  
QY 32 TCAGATACAGAAATCCGACATCAACTCCATGTTCTCGCGCGTAAGGAACCTTACCGTTG 91  
Db 114 TGAAGTTAAGTTGCGCAATTTCTACAGAGAGGAGGAGCTCACAGATTGCTGTTG 173  
QY 92 CTGAGAGGAATGCGAGGCTTTATGTTCTTGTGAGCGTTATTCGGCTTTAAGCCAT 151  
Db 174 CAGAGTATGAGATGCCAGGCTCTCATGAGTTGGCAAGGAATTCGACAGAGCAGCCTT 233  
QY 152 TGAAGGTTCTGAAATCTCTGGTTCCCTCCAGTACGATCCAGACAGCGCTCTCATCG 211  
Db 234 TGAAGGCGCCCGAATTGCTGTTCTATCCATGAGCGTCCAGACCGCGCTGCTTTG 293  
QY 212 AGACACTCAGAGCTTGTGCTGATGTCAGATGGGCTTCCTGCAACATCTTCTTACAC 271  
Db 294 AGACCTCACTGTTTGGGCGCTGAGGTTGTTGGCTTCTGCAACATTTCTCCACC 353  
QY 272 AAGNATACGCGCTGCTATTCGTTTCGGCCCAACAGGACACACAGGAGCCAGCGG 331

Db 354 AGATGAGGCTGACGCGCTATTCGTTGTCGSC---TCGSCACCGTGAAGAGCCAGCTG 410  
QY 332 GTATCCCAAGCTTTTCGCTGCTGAAGAGCGCAAAACACTCCAGATACCTGGGAGAAACATACC 391  
Db 411 GTGTTCCAGTATTTCGCGTGGAGAGGTGAGTCACTGGAGAGTACTGGTGGTGCATCAACC 470  
QY 392 GCGCTCTCATGCGCACGATGTCAGGCCCAAGAGGTCAGAGAGGTTCTGATGATGGTGTGATG 451  
Db 471 AGATCTTCAGCTGGGCGGATG---AGCTGCCAAACATGATCTCGACGACGCGGCTGACG 527  
QY 452 CTACACTCTCTCATCTCCAGAGGCTTCGAAATTCGAAACAGCGGCTCTGTTCCAGAGCCAA 511  
Db 528 CCACATGCGCTGTTATTCGCGTCCGAAATACAGACAGGCTGCTGTTTCCACCAGCAG 587  
QY 512 CAGAAGCTTGAACAACCTCGAATAACGCTGCGTCTTGTCTACACTCAAGCAGGTTCTTAACC 571  
Db 588 AGGCCAACGATTCGATGAGTACATCGCATTTTGGCGATGCTGGTGAGGTTCTTGCTG 647  
QY 572 AAGACAGAACCACTGCGACACAGTTGCTGCCGCAATGAACGCTGTTTCCGAGAGACAA 631  
Db 648 CAGAGCCTGGCAAGTGGGCAAGATTCGCTGAGGCGCTTAAGGCTGTCACCGAGAAACCA 707  
QY 632 CAACAGGTTGCCACCGCTCTACAGCTTCGAGAGAGGAGGCGCAAACTCTCTTCCAGCCA 691  
Db 708 CCACGCTGTCACCGCTGTACCACTTCGCTGAAGAGGCGTGTGCTTTCCAGCGA 767  
QY 692 TCACGCTCAACGACGCTGTTACAAAGTCCAAAGTTCGATTAACATCTACGCTGCGGCACT 751  
Db 768 TGAACGCTCAACGACGCTGTACCAAGTCAAGTTGATTAACAAGTACGCGCACCGCCACT 827  
QY 752 CCCTTATCGATGTATCAACCGTGTTCGATGTCCGATGTCGCGCGCAAGACACACTCTCG 811  
Db 828 CCCTGATCGACGCGCATCAACCGCCCACTGACATGCTCATGCGCGCGCAAGACGCTGTTG 887  
QY 812 TCATGGGTTACGCGGATGTCGCAAGGCTGCGCTCAATCCCTCCGCTGCGCAAGCGGCTC 871  
Db 888 TCTGCGGTTACGCGGATGTCGCAAGGCTGCGCTGAGGCTTTTCAGCGCGCAGCGGCTC 947  
QY 872 GCGTTATCATCAGAACTCGACCCCAATCTCGCTCTCCAGGCTGCCATGGAAGGCTACC 931  
Db 948 GCGTCAAGGTCACGGAAGCTGACCAATCAACGCTCTCTCAGGCTCTGTGATGATGGCTACT 1007  
QY 932 AGGTCGCGCGCATCGAGGAAGTGTCAAGAGTGTGATATCTTCGTTATCATGCAAGGAA 991  
Db 1008 CTGTGCTCACCGTTGATGAGGCCATCGAGAGCGCGACATCGTGTATCACCGGACCGCA 1067  
QY 992 ACTGCGATATCATCTCTGTTGACATGATGCGCCCAAGATGAGAGATGAGGCTATTGTCGGA 1051  
Db 1068 ACAGGACATCATTTCTTCGAGCAGATGCTCAAGATGAAGGATCACGCTCTGCTGGGCA 1127  
QY 1052 ACATCGGCGCTTCGATAACGAAATTTGATACAGATGCGCTCATGAAATACCCAGGCACTCA 1111  
Db 1128 ACATCGGTCATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1187  
QY 1112 AGCAGATCCCAATCAAGCCGCAATACGACATGTGGGAATTTCCAGATGCGCCACGCTATCC 1171  
Db 1188 CCGGACCAACGATCAAGCCACAGGTCGACGAGTTTCACTTTCTCCACCGCTCGCTCCATCA 1247  
QY 1172 TCCTTCTTGTGAGGCGCGCTTTTAACTTGTGCTGCGCTACAGGTACCGCTCTTTG 1231  
Db 1248 TCGTCTCTTCGAGGTCGCTGTTTGAACCTTTGGCAACCGCACCGGACACCATCATTTG 1307  
QY 1232 TTATGTCAATGTCTATTCACAAACACAGACACTCGCTCAGCTCGACCTCTACGAAAGAGAG 1291  
Db 1308 TCATGTCGAATCTTTTGGCGATCAGACCATTTGGCGAGATCGAATGTTTCAAAACGAAG 1367  
QY 1292 GAAATCTTCAGAGAAAGGTTTACACT 1319  
Db 1368 GACAGTACGAGAACGAGGTTCTACCGTCT 1395

RESULT 10  
US-09-925-301-217



```
; Sequence 217, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 217
; LENGTH: 2200
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2186)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-217

Query Match          22.7%; Score 331; DB 9; Length 2200;
Best Local Similarity 61.3%; Pred. No. 2.2e-97;
Matches 551; Conservative 0; Mismatches 345; Indels 3; Gaps 1;

QY 563 TCTTCAACCAAGACAAGAACCACTGGCACACAGATTGTCGGCGCATGAACGGTGTTCGCG 622
DB 480 TCACCAACTCATCCACACCAAGTACCOCGAGCTTCTGCCAGGCATCCGAGGCATCTCTG 539
QY 623 AAGACACACACAGGTGTCCACGGCTCTACAGCTCGAGAGAGGAGGCAACTCTCT 682
DB 540 AAGACACACAGCTGGGGTCCACAACTCTACAAGATGATGGCAATGGGATCTCAAGG 599
QY 683 TCCAGCCATCAAGTCAACGACGCTGTTCACAAAGTCCAGTTCGATAACATCAAGCT 742
DB 600 TGCCTGCATCAATGTCAATGATCTCGTCACACAGACAGATTGTGACACCTCTATGGCT 659
QY 743 GCGCGCACTCCCTTATCGATGGTATCAACCGTCTTCOGATGTCATGATCGCGGCAAGA 802
DB 660 GCGGGAGTCCCTCATAGATGGCATCAAGCGGCCACAGATGTGATGTCGCGCAAG 719
QY 803 CAGCTCTCGTCATGGGTTCAGCGATGTCCGCAAGGCTGGCTCAATCCCTCCGTGGCC 862
DB 720 TAGCCGTGTGACGAGGCTATGTTGATGTGGCAAGGCTGTGCCAGGCTCGCGGGTT 779
QY 863 AAGGCGCTCGGCTTATCATCACAAGAACTCGACCCCAATCTCGCTCTCCAGGCTGCCATGG 922
DB 780 TCGAGCCCGGCTCATATCATCCGAGATTGACCCCTCAACGACTCGAGGCTGCCATGG 839
QY 923 AAGGTACAGGTCCGCGCATCGAGGAAGTGTCAAGGATGTGATATCTTCGTTACAT 982
DB 840 AAGGTATAGGTGACCAACATCATGATGAGGCTGTGAGGAGGCAACATCTTTGTACCA 899
QY 983 GACAGGAACACTGGATATCATCTCTGTTGACATGATGGCCCGCAGATGAGGATAGGTA 1042
DB 900 CACAGGCTGTATTGACATCATCTCTGGCCGCACTTTGAGCAGATGAGGATGATGCCA 959
QY 1043 TTGTGGTAAACATCGGCCACTTTCGATAACGAAATTTGATACAGATGGGCTCATGAATACC 1102
DB 960 TTGTGTGTAACTTGGACATTTTGACGTGGAGATCGATGTCAGTGGCTCACAGGAACG 1019
QY 1103 CAGGCATCAAGCACATCCCAATCAACCCAGAAATACGACATGTGGGAATTCACAGATGGCC 1162
DB 1020 CCGTGAGAAAGGTGAAC---ATCAAGCCGAGGTGGACCGGTATCGGTTGAAGATGGGC 1076
QY 1163 AGCTATCTCTCTCTTCTGCTGAGGCGGCTTCTTAACCTTGGCTGGCTACAGTACAC 1222
DB 1077 GCGGCATCATCTCTGCTGGCGAGGGTGGCTGGTCAACCTGGTGTGTCATGGGCCACC 1136
QY 1223 CATCTTTCTGTTATGTCAATGTCTTTCACAAACACAGACACTCGCTCAGCTCGACCTCTACG 1282
```

```
DB 1137 CCAGCTTCGTGATGAGTAACTCTTTCCACCAACCAAGGTGATGGCGCAGATCGAGCTGTGGA 1196
QY 1283 AAAAGAGAGGAAATCTCGAAGAAAGGTTTACACACTTCGAAAGCATCTCGATGAAGAAG 1342
DB 1197 CCATCCAGACAGTACCCCGTTGGGTTTCATTTCTGCCCAAGAGCTGGATGAGCAG 1256
QY 1343 TCGCTCGCTTCCACCTCGGATCTTCGATGTCACACTTACAAAGTTTACAGAGCAGG 1402
DB 1257 TGGCTGAAGCCACCTGGGCAAGCTGAATGTGAAGTTGACCAAGCTTAACCTGAGAGCAAG 1316
QY 1403 CTGACTACATCAACGTTCCAGTTCAGGTGAGGCTCTTACAGTCTGATGCTTACCGTTAA 1461
DB 1317 CCCAGTACCTGGGCGATGTCTCTGTGAGGCCCCCTTCAAGCGGATACCTACCGTACTGA 1375

RESULT 11
US-10-044-090-344
; Sequence 344, Application US/10044090
; Publication No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bardman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 344
; LENGTH: 2429
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1468237CB1
US-10-044-090-344

Query Match          22.7%; Score 331; DB 13; Length 2429;
Best Local Similarity 61.3%; Pred. No. 2.3e-97;
Matches 551; Conservative 0; Mismatches 345; Indels 3; Gaps 1;

QY 563 TCTTCAACCAAGACAAGAACCACTGGCACACAGATTGTCGGCGCATGAACGGTGTTCGCG 622
DB 480 TCACCAACTCATCCACACCAAGTACCOCGAGCTTCTGCCAGGCATCCGAGGCATCTCTG 539
QY 623 AAGACACACACAGGTGTCCACCGCTCTTACAGCTCGAGAGGAGGCAAACTCTCT 682
DB 540 AAGACACACAGCTGGGGTCCACAACTCTACAAGATGATGGCAATGGGATCTCTCAAGG 599
QY 683 TCCAGCCATCAAGTCAACGACGCTGTTCACAAAGTCCAGTTCGATACATCTACGGCT 742
DB 600 TGCCTGCATCAATGTCAATGATCTCGTCACACAGACAGATTGTGACACCTCTATGGCT 659
QY 743 GCGCGCACTCCCTTATCGATGGTATCAACCGTCTTCOGATGTCATGATCGCGGCAAGA 802
DB 660 GCGGGAGTCCCTCATAGATGGCATCAAGCGGCCACAGATGTGATGATGTCGCGCAAG 719
QY 803 CAGCTCTCGTCATGGGTTCAGCGATGTTCGCAAGGCTGGCTCAATCCCTCCGTGGCC 862
DB 720 TAGCCGTGTGACGAGGCTATGTTGATGTGGCAAGGCTGTGCCAGGCTCGCGGGTT 779
QY 863 AAGGCGCTCGGCTTATCATCACAAGAACTCGACCCCAATCTCGCTCTCCAGGCTGCCATGG 922
DB 780 TCGAGCCCGGCTCATATCATCCGAGATTGACCCCTCAACGACTCGAGGCTGCCATGG 839
QY 923 AAGGTACAGGTCCGCGCATCGAGGAAGTGTCAAGGATGTGATATCTTCGTTACAT 982
DB 840 AAGGTATAGGTGACCAACATCATGATGAGGCTGTGAGGAGGCAACATCTTTGTACCA 899
QY 983 GACAGGAACACTGGATATCATCTCTGTTGACATGATGGCCCGCAGATGAGGATAGGTA 1042
DB 900 CACAGGCTGTATTGACATCATCTCTGGCCGCACTTTGAGCAGATGAGGATGATGCCA 959
QY 1043 TTGTGGTAAACATCGGCCACTTTCGATAACGAAATTTGATACAGATGGGCTCATGAATACC 1102
DB 960 TTGTGTGTAACTTGGACATTTTGACGTGGAGATCGATGTCAGTGGCTCACAGGAACG 1019
QY 1103 CAGGCATCAAGCACATCCCAATCAACCCAGAAATACGACATGTGGGAATTCACAGATGGCC 1162
DB 1020 CCGTGAGAAAGGTGAAC---ATCAAGCCGAGGTGGACCGGTATCGGTTGAAGATGGGC 1076
QY 1163 AGCTATCTCTCTCTTCTGCTGAGGCGGCTTCTTAACCTTGGCTGGCTACAGTACAC 1222
DB 1077 GCGGCATCATCTCTGCTGGCGAGGGTGGCTGGTCAACCTGGTGTGTCATGGGCCACC 1136
QY 1223 CATCTTTCTGTTATGTCAATGTCTTTCACAAACACAGACACTCGCTCAGCTCGACCTCTACG 1282
```



Db 178973 CTGTGGTTTGTAAACAGTTTGACAACTTGATGGTGGCCGCTCACTCTCTCCCTGATGTC 178914  
Qy 767 TGAACCGCTCTTCGATGTCATGATCGCGGCAAGACAGCTCTCGTCATGGTTACGGCG 826  
Db 178913 TCATCAGGCTACCGATGTTATGATGCTGGAAAGTGGCTGTTGTGGCTGATATGGTG 178854  
Qy 827 ATGTGCGCAAGGCTCGCTCAATCCCTCCGTGGCCAGGCTCGGTTATCATCAG 886  
Db 178853 ATGTTGGCAAGGTTGTCGCTGCAATGAAGCAGGCTGGTGTCTGTCATCGTACCG 178794  
Qy 887 AACTCGAACCAATCTGGCTCTCCAGGCTGCCAAGGCTACCAAGGCTCCCGCCATCG 946  
Db 178793 AGATTGATCCCATCTGTCGCTTCAGGCTCTCATGGAAGGCTTCAGGTTCTGACCTGG 178734  
Qy 947 AGAAGTGTCAAGATGTCGATATCTTCTGTTACATGCAAGAACTCGGATATCACT 1006  
Db 178733 AGGATGTTGTTCTGAGGCTGATATCTTGTCAACCAACCGGTAAAGAGCATCATCA 178674  
Qy 1007 CTGTTGACATGATGGCCAGATGAAGGATAGGCTATTGTCGTTAACTCGGCCACTTCG 1066  
Db 178673 TGGTTGACCATGAGGAAATGAAGAAATGCAATGCCATGTTTGAACATGTCATCTTG 178614  
Qy 1067 ATAAGAAATTGATACAGATGGCTCATGAATAACCAAGGCTCAAGACATCCCAATCA 1126  
Db 178613 ACAATGATCGACATGCTTGGCTGGAGAACTACCCGGGCTGAAGCGCATCAATCA 178554  
Qy 1127 AGCCAGATATGACATGTTGGAAATCCAGATGGCGAGC---TATCCTCTCTTCTGCTG 1183  
Db 178553 AGCCCCAACTGACAGATGGGCTTCCCTGAGACCAACCGGTATCATGTTCTGGCTG 178494  
Qy 1184 AGGGCGCTCTTAACTTGGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1243  
Db 178493 AGGTCGATTGATGATCTGGAGTGGCGGCTGACATGACATGCTGCTGCTGCTGCTGCT 178434  
Qy 1244 CATCAAAACAGACACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1297  
Db 178433 CCTTCAACAAACAGGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 178374  
Qy 1298 TCGAAGAGAGGTTTACACTTTCGAGACATCTCGATGAGAGTGGCTGCTGCTGCTGCTGCTGCT 1357  
Db 178373 AGAAGAGAGGTTTACGTTTCCCAAGCCTTGTATGAGAGGTTGGCTGCACTTACC 178314  
Qy 1358 TCGGATCTCTGATGCTCACTTACAAAGCTTACACAGAGCAGGCTGCTACATCAACG 1417  
Db 178313 TGGCAAACTTGGAGCTTAAGCTGACCCAGCTTAGCAAGTCCAGGCTGATTAATCACTG 178254  
Qy 1418 TTCCAGTTGAGGTCCTTACAACTCTGATGCTTACCGTTATTAA 1461  
Db 178253 TGCTGTTGAGGTTCCATACAAAGCTGCTCACTACAGGTACTAA 178210

RESULT 14  
US-10-342-224-51  
; Sequence 51, Application US/1034224  
; Publication No. US2003016229A1  
; GENERAL INFORMATION:  
; APPLICANT: Nathalie Verbruggen  
; TITLE OF INVENTION: Genes involved in Tolerance to Environmental Stress  
; FILE REFERENCE: CN-01208  
; CURRENT APPLICATION NUMBER: US/10/342,224  
; CURRENT FILING DATE: 2003-01-13  
; PRIOR APPLICATION NUMBER: US/09/762,154  
; PRIOR FILING DATE: 2002-02-02  
; PRIOR APPLICATION NUMBER: EP 98202634.6  
; PRIOR FILING DATE: 1998-08-04  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 51  
; LENGTH: 1398  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: CDS

; LOCATION: (1)....(1398)  
US-10-342-224-51  
Query Match 20.1%; Score 293.4; DB 12; Length 1398;  
Best Local Similarity 54.2%; Pred. No. 3.7e-85;  
Matches 715; Conservative 0; Mismatches 556; Indels 48; Gaps 4;  
Qy 146 AGCCATGGAAGGTTGACAAATCTCTGTTCCCTCCACATGACAGTCCAGAGCGGCTCC 205  
Db 47 AGCCATGTTAGAGGCGGTTCAATCACTGGATCTCTACCTGACCGCTCGACCGCGTATCC 106  
Qy 206 TCATCGAGACATCAACAGTCTTGTGTCGATGTCAGATGGCTTCTTCCGCAACATCTTCT 265  
Db 107 TTATCGGAACCTTACCGCTTTAGCGGTGGAGTTAGATGTGGCTCTTGCCTCAGTGTG 166  
Qy 266 TTACAAGATACAGCGGCTGCTGCTATCTGTTGTGGCCCCAACAGGCACACAGAGAAGC 325  
Db 167 ACCGTTGGGACACGCAAGTGGCGCATAGC-----GC 199  
Qy 326 CAGCCGTTATCCAGTCTTGGCTGGAAGGGCGAAACACTCCAGAAATCTGGGAGACA 385  
Db 200 GTGATAGCTGTGGCGTGTTCGATGGAAGCGAGGTACGCGACAAGAGTACTGGTCTCGA 259  
Qy 386 CATACCGGCTCTTCACATGCGCCAGATGTTCAAGGCCCCACAGCAGGTTGTCGATGATG 445  
Db 260 CTGAACCGACCTTGACTGGGCGCCCGTGGTGACCCGACTTCGATGTCGATGATG 319  
Qy 446 GTGATGTCATCT 493  
Db 320 GTGCGATTCGCTTTTGAATCAAGATGCGTAAAGCTGGGAGGAAATATGCTAAATCTG 379  
Qy 494 GTGCTGTTCCAGAGCCCAAGAGCTGACACCTCGAATACCTGCTGCTCTCTCTCTCTCT 553  
Db 380 GAAAGTTCCAGATCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 439  
Qy 554 TCACAGAGGCTCTTCAACCAAGCAAGAACCTCTGCGCACACAGTGTGTCGGCGCATGAACG 613  
Db 440 TTAGGAGGCGTTGAAATCTGATCTTTTAAATACATAGACGACCCAGTTGCTCTGTTG 499  
Qy 614 GTGTTTCGAAGAGACAACAACAGGTGTCACCGCTCTACAGCTCGAGAGAGAGGCA 673  
Db 500 GTGTTTCTGAGAACTACTACTGCTGTAAAGAGAGTTACCAAAATGACGCGCAATGTA 559  
Qy 674 AACTCTCTTCCAGCCATCAAGTCAAGAGCTGTTACAAAGTCCAGTTCAGTAACA 733  
Db 560 CTTTCTCTTTCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 619  
Qy 734 TCTAGGCTGCGCCACT 793  
Db 620 TGTACGATGCGCCACT 679  
Qy 794 GCGCAAGACAGTCT 853  
Db 680 CTTGCAAGGTAGCGCTTGTAGGCGGTTACGAGATGCTTTTAAAGGTTGGGTTGCTGCT 739  
Qy 854 TCCGTGGCCAAAGCGCTCGGTTATCATCACAGAACTCGACCCCAATCTGCGCTCTCAGG 913  
Db 740 TGAAGCAAGCTGTTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 799  
Qy 914 CTGCCATGGAAGGCTACAGGCTCCGCGCATCGAGGAAGTCTGCAAGGATGTCGATATCT 973  
Db 800 CTACCATGGAAGGTAGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 859  
Qy 974 TCGTTACATGCAAGAACTCGGATATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1033  
Db 860 TCGTTACGACAAACCGGTAAAGAGGCTCTCATGTTGACCATGAGGCGGAATGAAGA 919  
Qy 1034 ATAGGCTATTGTGCGTAAACATCGCCACTTTCGATAAGAAATTCGATACAGATGCTCA 1093  
Db 920 ACCAGGCCATGTTTGCACATCTCGACCTTTTCGACATGAAATTCGACATGCGGCTCT 979  
Qy 1094 TGAATACCCAGGCTCAAGCACTCCCAATCAAGCCAGATACGACATGTGGGAATCC 1153

Db 980 AGACCTTCCTCGAGTGAACGGATCAATCAAGGCCACAGTACAGATGGTCTTTC 1039  
Qy 1154 CAGATGGCCAC---GCTATCCTCTTCTTGTGTGAGGGCCGCCCTTTTAACCTTGGCTGCG 1210  
Db 1040 GCGACACCAACAGAGGTATCATTTGTCAGCGAGGGGGCGTCTCATGATGGATGCG 1099  
Qy 1211 CTACAGGTACCCATCTTCTGTTATGTCAATGTCAATTCACAAACAGACACTCGCTCAGC 1270  
Db 1100 CCACTGGACACCCAGCTTCGGAGCTCTGCTCTTTCACTAACCAAGTCAGTCTCAGC 1159  
Qy 1271 TCGACCTCT-----ACGAAAGAGAGGAAATCTCGAAGAGAGGTTTACACACTTCCGA 1324  
Db 1160 TCGAGTCTGCGGGAGAGACACCGCAAGTATGAGAGAAAGTGTACGCTTCCAA 1219  
Qy 1325 AGCATCTCGATGAAGAAGTCTGCTCTCCACCTCGGATCTCTCGATGTCACCTTACAA 1384  
Db 1220 AGCACTTTGAGAAGAGGTTGCGGCCCTTCATCTCTGTAAGCTCGGAGCAAGGCTCACTA 1279  
Qy 1385 AGCTTACACAGACAGGCTGACTACATCAAGTTCCAGTTGAGGTCCTTACAGTCT 1443  
Db 1280 AGCTTAGTGTGACACCTTCTTTGACGAGACCCAGTTGAAGTCTGTAAGAGCCT 1338

RESULT 15

US-09-738-626-838/c  
; Sequence 838, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 838  
; LENGTH: 720  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-838

Query Match 19.7%; Score 288.4; DB 10; Length 720;  
Best Local Similarity 64.6%; Pred.No.1.1e-83;  
Matches 430; Conservative 0; Mismatches 236; Indels 0; Gaps 0;  
Qy 421 CCACAGCAGGTTGTGCGATGATGGTGGTATGCTACACTCTCATCTCCAAAGGGCTTCGAA 480  
Db 666 CCAACATGATCCTCGACGAGCGGTGACGCCACCATGGCTGTATTTCGGGTGCGAA 607  
Qy 481 TTCGAACAGCGGTGCTGTTCAGAGCCAAACAGAGCTGACAACTCGAATACCGTTCG 540  
Db 606 TAGCAGCAGGCTGTCTGTGTTCCACAGCAGAGGCCAACGATTCCGATGATACATCGCA 547  
Qy 541 GTTCTGTGTACATCAGCAGGCTTCAACACAGACAGACACCTGGCAGCAGTTGCT 600  
Db 546 TCTTGGGCGATGCTGCGTGGGTTCTTGTGTCAGAGCCTGGCAAGTGGGGCAAGATCGCT 487  
Qy 601 GCGGCGATGAACGGTGTTCGGAAGAGACAAACAGGTGTCCACCGCTCTTACCAGCTC 660

Db 486 GAGGCGGTTAAGGGTGTCTACCGAGAAACACACCGGTGTGACCGCTGTACACACTTC 427  
Qy 661 GAGAAGGAGGGGCAAACTCCTCTTTCCAGCATCAACGTCAACGACGCTGTACAAAGTCC 720  
Db 426 GCTGAGAGGCGTGTGCTTTTCCAGCGATGAAGTCAACGACGCTGTACCAAGTCC 367  
Qy 721 AAGTTGATTAACATCTACCGGTGCGGCCACTCCCTTATCGATGGTATCAACGGTGTCTTC 780  
Db 366 AAGTTTATTAACAGTACCGCACCGGCCACTCCCTGATCGAGCGCATCAACCGGCGCACT 307  
Qy 781 GATGTATGATTCGCGCGCAAGACAGCTCTCTGCTATGTTTACGGGATGTGCGCAAGGC 840  
Db 306 GCATCTCTATGCGCGGCAAGACGTTGTTCTGCGGTTACGGGATGTGCGCAAGGC 247  
Qy 841 TCGGCTCAATCCCTCCGTCGCGCAAGCGCTCGGTTATCATCACAGAACTCGACCCCAATC 900  
Db 246 TCGGCTGAGGCTTTTGAAGCGCCAGGCGCTCGCTCAAGGTCAACGAGCTGACCCCAATC 187  
Qy 901 TCGGCTCTCCAGGCTGCCATGGAAGCTACAGGTTCGCGCGCATCGAGGAAGTGTCTAAG 960  
Db 186 AACGCTCTTCAGGCTCTGATGGATGGCTACTCTGTGGTCAACGTTGATGAGGCGCATCGAG 127  
Qy 961 GATGTGATATCTTCTGTATGTCACAGGAACTCGGATATCATCTCTGTGACATGATG 1020  
Db 126 GACGCCGACATCTGTATCAACCGGACCGGCAACAGGACATCATTTCTTCGAGCAGATG 67  
Qy 1021 GCGCAGATGAAGGATAAGGCTATTGTGGTAACATCGGCGCACCTTCGATAACGAAATTCAT 1080  
Db 66 CTCAGATGAAGGATCACGCTCTGCTGGGCAACATCGGTCACTTTGATAATGAGATCGAT 7  
Qy 1081 ACAGAT 1086  
Db 6 ATGCAT 1

Search completed: October 18, 2003, 00:31:06  
Job time : 455 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame plus n2p model

Run on: October 15, 2003, 14:41:49 ; Search time 59 Seconds  
(without alignments)

Title: US-09-759-990A-1  
 Perfect score: 2737  
 Sequence: 1 atgcttggaaaatcacctac.....ctgatgcttaccgttattaa 1461

Scoring table:	BLOSUM62	
	Xgapop 10.0	Xgapext 0.5
	Ygapop 10.0	Ygapext 0.5
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 566616

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

## Listing first 45 summaries

Command line parameters:

```
-Q/-cseq=2.1/USFO-spool/US09759990/runat_15102003_115153_2388/app_query.fasta.1.1607
-QC/frame+ n2p.model -DEV=xlp
-DB-P1R 76 -QWTS-fastan -SUFFIX=rpr -MINMATCH=0.1 -LOCPCL=0 -LOCALTEXT=0
-UNITS=bits -START=1 -END=1 -WATXr-blosum62 -TRANS-human40.cdi -LIST=45
-DCCALING=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT-plo -NORM=ex -HEADSPZ=570 -MINLEN=0 -MAXLEN=2000000000
-USER=US09759990@ccn.1.1.77@runat_15102003_115153_2388 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DESPLOCK=100 -LONGLOG
-FWT TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=-7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=-6 -DELEXT=-7
```

```
Database : PIR_76.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	1550.5	56.6	495	2	B70593	adenosylhomocystei
2	1537.5	56.2	492	2	D87005	probable S-adenosyl
3	1537	56.2	485	2	T71400	adenosylhomocystei
4	1524	55.7	485	2	T06764	adenosylhomocystei
5	1520	55.5	485	2	S38379	adenosylhomocystei
6	1509	55.1	485	2	S71621	adenosylhomocystei
7	1406	51.4	446	2	D82730	adenosylhomocystei
8	1382	50.5	466	2	AF2580	S-adenosylhomocyst
9	1382	50.5	466	2	D97362	adenosylhomocystei
10	1374	50.2	463	2	H87280	adenosylhomocystei
11	1352.5	49.4	462	1	A46035	adenosylhomocystei
12	1343	49.1	481	2	AG3505	adenosylhomocystei
13	1332	48.7	479	1	A54040	adenosylhomocystei
14	1289	47.1	432	1	A26583	adenosylhomocystei

## ALIGNMENTS

## RESULT 1

B70593

adenosylhomocysteine (EC 3.3.1.1) - Mycobacterium tuberculosis (strain H37Rv)

C:Species: *Mycobacterium tuberculosis*

C;Date: 17-Jul-1998 #sequence revision 17-Jul-1998 #text change 20-Jun-2000

C;Accession: B70593

R; Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, J.; Connor, R.; Davies, R.; DeLam, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajadream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of *Mycobacterium tuberculosis*

A;Reference number:

A;Accession: B70593

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-495 <COL>

A; Cross-References: GB:295121; GB:A

A; Experiment 2  
C. Genetics;A:Gene. cashu  
C;GENETICS:

H; gene: *sahH*  
C-Superfamily: *adenosyl homocysteine*

C: Keywords: NAD: thioether hydrolysis

Enzymes: NAD, succinate dehydrogenase  
 F:276-305/Region: beta-alpha-beta nucleotide-binding fold

Alignment Scores:		
Pred. No.:	2,650-94	455
Score:	1550.50	237
Length:		237
Matches:		68
Conservative:	62.2%	109
Best Local Similarity:	76.5%	
Mismatches:	56.6%	3
Query Match:		3
DB:	2	
Gaps:		

US-09-759-990A-1 (1-1461) X B70593 (1-495)

34 GAGTACAGAAATTGCGGACATCAACCTCCATCTGTTCTCCGCGCTACGCAACCTTACCCCTTCCT 93

[illegible]

Db  
20 AspPheIysIleAlaAspLeuSerLeuAlaAspPheGlyArgLysGlnLeuArgIleAla 39

THE UNIVERSITY OF CHICAGO

QY 94 GAGAGGAAGATGCCAGGCTCTTATGGTTCTTTCGTGAGCGTTATTCGCCGCTCTCTAAGCCATG 153  
Db 40 GluHisGluMetProGlyLeuMetSerLeuArgArgGluTyrAlaGluValGlnProLeu 59  
QY 154 AAGGGTCTCAGAAATCTCTGGTTTCCTCCATGACAGTCCAGACAGCGCTCTCATCGAG 213  
Db 60 LysGlyAlaArgLysSerGlySerLeuHisMetThrValGlnThrAlaValLeuLeuGlu 79  
QY 214 ACACTCACAGCTCTTGGTGTCTGATGTACAGATGGCTTCTCTGACACATCTCTCTACAA 273  
Db 80 ThrLeuThrAlaLeuGlyAlaGluValArgTyrAlaSerCysAsnIlePheSerThrGln 99  
QY 274 GATACAGCCGCTGCTGCTATCGTTGTTCGGCCCAACAGGCACACAGAGAGCAGCCCGGT 333  
Db 100 AspHisAlaAlaAlaValValGlyProHisGlyThrProAspGluProLysGly 119  
QY 334 ATCCCACTCTTCGCTGGAAGGGCGAAACACATCCAGAAATCTGGAGAACACATACCGC 393  
Db 120 ValProValPheAlaTyrLysGlyGluThrLeuGluLysTyrTrpAlaAlaGluGln 139  
QY 394 GCTCTCATGTCGCAGATGGTCAAGGCCCA---CAGCAGGTTGTGCATGATGGTGTGAT 450  
Db 140 MetLeuThrTyrProAspProAspLysProAlaAsnMetIleLeuAspAspGlyAsp 159  
QY 451 GCTACACTCTCATCTCCAGGGCTTCGAATTCGAAACAGCCGGTCTGTTCCAGAGCCA 510  
Db 160 AlaThrMetLeuValLeuArgGlyMetGlnTyrGluLysAlaGlyValValProProAla 179  
QY 511 ACAGAAGCTCAACAACCTCGAATCCGCTGCTTGTCTACATCAAGCAGGTCCTTCAAC 570  
Db 180 GluGluAspAspProAlaGluTyrLysValPheLeuAsnLeuLeuArgThrArgPheGlu 199  
QY 571 CAAGAAGAAGCAACATGGCCACACAGTGTCTGCCGCATCAAGCGTGTTCGAGAGACA 630  
Db 200 ThrAspLysAspLysTyrThrLysIleAlaGluSerValLysGlyValThrGluGluThr 219  
QY 631 ACAACAGGTGTCCACGGCTCTACCACTCGAGAGGAGGCGCAAACTCCTCTTCCAGCC 690  
Db 220 ThrThrGlyValLeuArgLysTyrGlnPheAlaAlaGlyAspLeuAlaPheProAla 239  
QY 691 ATCAACGCTCAACAGCGCTGTACAAAGTCCCAAGTTCGATAACATCATCCGCTGCCGCC 750  
Db 240 IleAsnValAsnAspSerValThrLysSerLysPheAspAsnLysTyrGlyThrArgHis 259  
QY 751 TCCCTTATCATGATGATCAACCGGTCTTCGATGTTCATGTCGCGCGCAGACAGCTCTC 810  
Db 260 SerLeuIleAspGlyIleAsnArgGlyThrAspAlaLeuLysGlyLysLysValLeu 279  
QY 811 GTCATGGTTTACGGGATGTCGCAAGGGCTGCGTCAATCCCTCGTGGCCCAAGCGCT 870  
Db 280 IleCysGlyTyrGlyAspValGlyLysGlyCysAlaGluAlaMetLysGlyGlnGlyAla 299  
QY 871 CGCGTTATCATCACAAGAACTCGACCCAAATCGCGCTCTCCAGGTCGCATGAAGCGCTAC 930  
Db 300 ArgValSerValThrGluLeAspProIleAsnAlaLeuGlnAlaMetMetGluGlyPhe 319  
QY 931 CAGTCTCCGCGCATCGAGAGATCGTCAAGGATGTCGATATCTTCGTGTATCATGCACAGA 990  
Db 320 AspValValThrValGluGluAlaIleGlyAspAlaAspIleValValThrAlaThrGly 339  
QY 991 AACTGCGCATATCATCTCTGTGACATCATGTCGCGCAGATGAAGGTAAGCTATTCGCT 1050  
Db 340 AsnLysAspIleIleMetLeuGluHisIleLysAlaMetLysAspHisAlaLeuLeuGly 359  
QY 1051 AACATCGGCCACTTCGATAACGAATATGATACAGATGCGCTCATGATAATACCCAGGCATC 1110  
Db 360 AsnIleGlyHisPheAspAsnGluIleAspMetAlaGlyLeuGluArg--SerGlyAla 378  
QY 1111 AAGCACATCCCAATCAGCCAGAAATACGATGTGGAAATTCGCCAGAT---GGCCACGCT 1167  
Db 379 ThrArgValAsnValLysProGlnValAspLeuThrThrPheGlyAspThrGlyArgSer 398

QY	1168	ATCCTCTCTTCTGCTAGGGCGGCTCTTTAACTTGGCTGGCTACAGTCAACCCACT	1227
Db	399	llelleValLeuSerGluGlyArgLeuLeuAsnLeuGlyAsnAlaIarThrGlyHisProSer	418
QY	1228	TTCCGTATGTCAATGTTCATTACAAAACACACACTCGCTCAGCTTCAGCAAG	1287
Db	419	PheValMetSerAsnSerPheAlaAsnGlnThrIleAlaGlnIleGluLeuTrpThrLys	438
QY	1288	AGAGGAATCTCGAGAGAGAGGTTTACACTTCCGAAGCATCTCGATGAGAGAGTCCGT	1347
Db	439	AsnAspGlnTrpAspAsnGluValTrpArgLeuProLysHisLeuAspGluLysValala	458
QY	1348	CGCTCCCACTCGGAPCTCTCGATGCTCCACTTACAAAAGCTTACACAGACAGCGGTGAC	1407
Db	459	ArgIleHisValGluAlaLeuGlyGlyHisLeuThrLysLeuThrLysGluGlnAlaGlu	478
QY	1408	TACATCAACCTCCAGTGTGAGGTCCTTCAAGTCTGATGCTTACCGCTTAT	1458
Db	479	TyrLeuGlyValAspValGluGlyProTyrLysProAspHisTyrArgTyr	495
RESULT 2			
D87005			
probable S-adenosyl-L-homocysteine hydrolase [imported] - Mycobacterium leprae			
C:Species: Mycobacterium leprae			
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001			
C:Accession: D87005			
R: Cole, S.T.; Biglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, B.; Davies, R.M.; Devlin, K.; Duthey, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; cam, M.A.; Rutherford, K.M.			
Nature 409, 1007-1011, 2001			
A: Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Square			
A: Title: Massive gene decay in the leprosy bacillus.			
A: Reference number: A86909; PMID:21128732; PMID:11234002			
A: Accession: D87005			
A: Status: Preliminary			
A: Molecule type: DNA			
A: Residues: 1-492 <STO>			
A: Cross-references: GB:AL450380; NID:gl3092881; PIDN:CAC30280.1; GSPDB:GN00147			
C: Genetics:			
A: Gene: sahH			
C: Superfamily: adenosylhomocysteinase			
Alignment Scores:			
Pred. No.: 1,9e-93 Length: 492			
Score: 1537.50 Matches: 297			
Percent Similarity: 76.52% Conservative: 68			
Best Local Similarity: 62.26% Mismatches: 109			
Query Match: 56.17% Indels: 3			
DB: 2 Gaps: 3			
US-09-759-990A-1 (1-1461) x D87005 (1-492)			
QY	34	GAGTACAGAAATTCGCCACATCAAACTCCATGTTCTTCGGCCGCTAAGAACTTACCTTCTCT	93
Db	17	AspPheLysValAlaAlaAspLeuSerLeuAlaAsnPheGlyArgLysGluLeuAspLeuAla	36
QY	94	GAGAGGAATGCCAGGTCCTTATGGTTCTTCGFGAGCGCTTATCCGCTTCTAAGCATTTG	153
Db	37	GluTyrGluMetProGlyLeuMetSerLeuArgHisGluTyrAlaGluValGlnProLeu	56
QY	154	AAGGGTGTCAAGATCTCTGGTTCCCTCCACATGACAGTCCAGACAGCGCTCTCATCCAG	213
Db	57	LysGlyAlaArgIleSerGlySerLeuHisMetThrValGlnThrAlaValLeuIleGlu	76
QY	214	ACATCTCACGCTTGTGGTGTGATGCTCAGATGGCTTCTCTCAACATCTTCTCTACAA	273
Db	77	ThrLeuThrAlaLeuGlyAlaGluValArgTrpAlaSerCysAsnIlePheSerThrGln	96
QY	274	GATACACCGCTCTGCTCTATGGTTCTCGGCCCAACAGGACACACAGAGACGCCCGGT	333
Db	97	AspHisAlaAlaAlaValAlaValGlyProTyrGlyThrProGluGluProLysGly	116
QY	334	ATCCACGCTCTTCCTCGAAGGCCGAAACACTCCACAGATATCTGGGAGAACACATACCG	393



QY 319 GAGAACCCAGCCGGTATCCAGTCTTCGCTGGAGGCGCAACATCCACGAATACCTGG 378  
Db 106 -----ValPheAlaTrpIysGlyGluThrLeuGlnGluTrp 118  
QY 379 GAGAACATACCCGGCTCTCATCGCCAGATGTCAGGCCACAGCAGGTCTCGAT 438  
Db 119 TrpCysThrGluArgAlaLeuAspTrpGlyProGlyGlyProAspLeuIleValAsp 138  
QY 439 GATGGTGTGATGTACACTCTCATCTCCAGGCGCTTCGAA-----TTCGAA 486  
Db 139 AspGlyGlyAspAlaThrLeuLeuIleHisGluGlyValIysAlaGluGluIlePheGlu 158  
QY 487 ACAGCCGGTGTCTTCAGAGCCACAGAAAGCTGACACCTCGAAATACCGCTGCTTCT 546  
Db 159 LysThrGlyGlnValProAspProThrSerThrAspAsnProGluPheGlnIleValLeu 178  
QY 547 GCTACACTCAAGCGTCTTCCACCAAGACAAAGACCACTGGCACACAGTCTCCCGGC 606  
Db 179 SerIleIleLysGluGlyLeuGlnValAspProLysLysTyrHisLysMetLysGluArg 198  
QY 607 ATGAACGGTGTTCGAGAGACAAACAGAGTGTCCACCGCTCTACCAAGTCCGAGAG 666  
Db 199 LeuValGlyValSerGluGluThrThrThrGlyValIysArgLeuTyrGlnMetGln 218  
QY 667 GAGGCAAACTCTCTCCAGCCATCAAGTCACAGCGCTTACAAAGTCCAGTTC 726  
Db 219 AsnGlyThrLeuLeuPheProAlaIleAsnValAsnAspSerValThrLysSerLysPhe 238  
QY 727 GATAACACTAGCGTGTCCGCCACTCTCTTATCGATGTATCAACCGTCTCCGATGTC 786  
Db 239 AspAsnLeuTyrGlyCysArgHisSerLeuProAspGlyLeuMetArgAlaThrAspVal 258  
QY 787 ATGATCGCGGAGACAGACTCTCGTCATGCTGATGGCGATGTCGCGAAGGCTCGCCT 846  
Db 259 MetIleAlaGlyLysValAlaValIleCysGlyTyrGlyAspValGlyLysGlyCysAla 278  
QY 847 CAATCCCTCCGGCGCAGGCGTCCGCTTATCATCAGAACTCCAGCAACTCCGCAATCTCGCCT 906  
Db 279 AlaAlaMetLysThrAlaGlyAlaArgValIleValThrGluIleAspProIleCysAla 298  
QY 907 CTCAGGCTCCATGGAAGGCTACAGGTCCCGCCGATCGAGAGTGTCTCAAGATGTC 966  
Db 299 LeuGlnAlaLeuMetGluGlyLeuGlnValLeuThrLeuGluAspValValSerGluAla 318  
QY 967 GATATCTCTGTACATGACAGGAACTCGGATCATCTCTGTGACATGATGCCAG 1026  
Db 319 AspIlePheValThrThrThrGlyAsnLysAspIleIleValMetValAspHisMetArgLys 338  
QY 1027 ATGAAGGATAAGGCTATTGTCTGTAACATCGGCCACTTCCGATAACGAAATTCATCAGAT 1086  
Db 339 MetLysAsnAsnAlaIleValCysAsnIleGlyHisPheAspAsnGluIleAspMetLeu 358  
QY 1087 GGCTCATGAATACCCAGGATCAGACATCCCAATCAAGCCAGAAATACGATGTGG 1146  
Db 359 GlyLeuGluThrTyrProGlyValIysArgIleThrIleLysProGlnThrAspArgTrp 378  
QY 1147 GAATCCCAAGGCGCCAGCT---ATCCTCTCTTGTGTGAGGCGCCCTTCTTAACCTT 1203  
Db 379 ValPheProGluThrLysAlaGlyIleIleValLeuAlaGluGlyArgLeuMetAsnLeu 398  
QY 1204 GGCTGCGCTACAGGTACCCATCTTCTGTTATGTCAATGTCAATCAACACGACACTC 1263  
Db 399 GlyCysAlaThrGlyHisProSerPheValMetSerCysSerPheThrAsnGlnValIle 418  
QY 1264 GCTCAGCTCGACTCTACGAAAGAGA-----GGAATCTCGAGAGAGGTTTACACA 1317  
Db 419 AlaGlnLeuGluLeuTrpAsnGluLysAlaSerGlyLysTyrGluLysLysValTyrVal 438  
QY 1318 CTTCCGAAGCATCTCGATGAAGAAGTCTCGCTCCGCTCCACCTCGGATCTCTCGATGTCCAC 1377  
Db 439 LeuProLysHisLeuAspGluLysValAlaLeuLeuHisLeuGlyLysLeuGlyAlaArg 458  
QY 1378 CTTACAAAGCTTACAGAGCAGGCTGACTACATCAACGTTCCAGTTGAGGCTCTTAC 1437

Db 459 LeuThrLysLeuSerLysAspGlnSerAspTyrValSerIleProIleGluGlyProTyr 478  
QY 1438 AAGTCTGATCTTACCGTTAT 1458  
Db 479 LysProProHisTyrArgTyr 485  
RESULT 4  
T06764  
adenosylhomocysteine (EC 3.3.1.1) - wheat  
N:Alternate names: S-adenosyl-L-homocysteine hydrolase  
C:Species: Triticum aestivum (common wheat)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 23-Jul-1999  
C:Accession: T06764  
R:Richards, K.D.; Gardner, R.C.  
A:Description: The influence of aluminium on histone, heat shock and S-adenosyl-L-homocysteine  
A:Reference number: 215795  
A:Accession: T06764  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-485 <RIC>  
A:Cross-references: EMBL:L11872; NID:gi70772; PIDN:AAA34303.1; PID:gi70773  
C:Genetics:  
A:Gene: SH6.2  
C:Superfamily: adenosylhomocysteine  
C:Keywords: NAD; thioether hydrolase  
F:263-292/Region: beta-alpha-beta NAD nucleotide-binding fold  
F:86,120/Active site: Cys #status predicted  
Alignment Scores:  
Pred. No.: 1,46e-92 Length: 485  
Score: 1524.00 Matches: 301  
Percent Similarity: 73.51% Conservative: 57  
Best Local Similarity: 11.81% Mismatches: 113  
Query Match: 55.68% Indels: 16  
DB: 2 Gaps: 4  
US-09-759-990A-1 (1-1461) x T06764 (1-485)  
QY 19 ACTGTGTCCTCATTCGATACAGAAATCCGACATCACTCCATGTTCTCGGCGGTAAG 78  
Db 8 ThrSerSerGlyArgGluTyrLysValIysAspLeuPheGlnAlaAspPheGlyArgLeu 27  
QY 79 GAATCTACCTCTCTCAGAGGAATGCCAGCTTATGTTCTTCGTGAGGCTATTC 138  
Db 28 GluLeuGluLeuAlaGluValGluMetProGlyLeuMetAlaCysArgThrGluPheGly 47  
QY 139 GCTTCTTAAGCATTGAAGGCTGTCTGATCTCTGTTTCCCTCCATCAGATCCAGACA 198  
Db 48 ProSerGlnProPheLysGlyAlaArgIleSerGlySerLeuHisMetThrIleGlnThr 67  
QY 199 GCGTCTCATTCGAGACATCTCAGCTCTTGGTGTGATGTCAGATGGGCTTCTGCAAC 258  
Db 68 AlaValLeuIleGluThrLeuThrAlaLeuGluValArgTrpCysSerCysAsn 87  
QY 259 ATCTTCTCTACACAGATACAGCGCTCTGCTATCTGTTCTCGGCCAACAGGCACCA 318  
Db 88 IlePheSerSerGlnAspHisAlaAlaAlaIleAlaArgAspSerAlaAla----- 105  
QY 319 GAGAACCCAGCCGGTATCCAGTCTTCACATGGCCAGATGGTCAAGGCCACAGCAGGTGTCTGAT 438  
Db 119 TrpCysThrGluArgCysLeuAspTrpGlyValGlyGlyProAspLeuIleValAsp 138  
QY 439 GATGGTGTGATGTACACTCTCATCTCCAGGCGCTTC-----CAATTCGAA 486  
Db 139 AspGlyGlyAspAlaThrLeuLeuIleHisGluGlyValIysAlaGluGluPheGlu 158  
QY 487 ACAGCCGGTGTCTTCAGAGCCACAGAAAGCTGACACCTCGAAATACCGTCTCCGAT 546



```

Db      159 LysSerGlyLysValProAspProGluSerThrAspAsnProGluPheLysLeuValLeu 178
Qy      547 GCTACACCAAGCAGGCTTCAACCAAGACAGACCACTGGCAGACAGTCTGCGCGG 606
Db      179 ThrIleIleArgAspGlyLeuLysThrAspAlaSerLysTyArgLysMetLysGluArg 198
Qy      607 ATGACAGGTGTTCCGAGAGACACACAGGTGTCACCGCTCTACCGAGTTCGAGAG 666
Db      199 LeuValGlyValSerGluGluThrThrThrGlyValLysArgLeuTyrgLmMetGlnGlu 218
Qy      667 GAGGCAAACTCCTCTCCAGCCATCAAGTCAAGCAGCTGTACAAAGTCAAGTTC 726
Db      219 SerGlyThrLeuLeuPheProAlaIleAsnValAsnAspSerValThrLysSerLysPhe 238
Qy      727 GATAACATCTACGCTCGCCCACTCCCTTATCGATGATATCAACCGCTCTCCGATGTC 786
Db      239 AspAsnLeuTyrgLysCysArgHisSerLeuProAspGlyLeuMetArgAlaThrAspVal 258
Qy      787 ATGATCGGCGGACAGACAGCTCTCGTCATGGGTACGGCATGTGCGGAGGCTGGCT 846
Db      259 MetIleAlaGlyLysValAlaValCysGlyTyrgLysAspValGlyLysGlyCysAla 278
Qy      847 CAATCCCTCCGTGGCGCAAGCGCTCGCTTATCATCATCAGAACTCGAACCAATCTGGCT 906
Db      279 AlaAlaLeuLysGlnAlaGlyAlaArgValIleValThrGluLeuAspProIleCysAla 298
Qy      907 CTCAGGCTCGCATGAGAGCTACACAGTCCGCGCATTCGAGGAATGCTCAGAGATGC 966
Db      299 LeuGlnAlaLeuMetGluGlyIleGlnLeuThrLeuGluAspValValSerGluAla 318
Qy      967 GATATCTCGTATCATGCCAGAACTCGCATATCATCTCTGTTACATGATGCCCGAC 1026
Db      319 AspIlePheValThrThrThrGlyAsnLysAspIleLeuMetValAspHisMetArgLys 338
Qy      1027 ATGAGGATGAAGGCTATTGTCGTAATACATCGGCCCACTTCGATAACGAAATTCATACAGAT 1086
Db      339 MetLysAsnAsnAlaIleValCysAsnIleGlyHisPheAspAsnGluLeuAspMetAsn 358
Qy      1087 GGCCTCATGAATACCCAGGCATCAGACATCCCATCCAGGAGGAGGAGATGAGTGG 1146
Db      359 GlyLeuGluThrTyrgLysValLysArgIleThrIleLysProGlnThrAspArgTrp 378
Qy      1147 GAATTCGCCAGATGCCAC---GCTATCTCTCTCTGCTGAGGCGCTCTTAACTT 1203
Db      379 ValPheProGluThrLysThrGlyIleIleValLeuAlaGluGlyArgLeuMetAsnLeu 398
Qy      1204 GCTCGGTACAGTCCACCATCTTTCGTTATGTCAATGTCAATCAACACACACACTC 1263
Db      399 GlyCysAlaThrGlyHisProSerPheValMetSerCysSerPheThrAsnGlnValIle 418
Qy      1264 GCTCAGCTCGACCTCTACGAAAGAGA-----GGAATCTCAGAGAGAGTTCACAC 1317
Db      419 AlaGlnLeuGluLeuTrpAsnGluLysAlaSerGlyLysTyrgLysLysValTyrgVal 438
Qy      1318 CTTCCGAGCATCTCGATGAGAGTCGCTCGCTCCACCTCGCATCTCTCGATGTCAC 1377
Db      439 LeuProLysHisLeuAspGluLysValAlaAlaLeuHisLeuGlyLysLeuGlyAlaArg 458
Qy      1378 CTTCAAAAGCTTACAGAGAGAGGCTGATCATATCAACGTTCCAGTTGAGGCTCTTAC 1437
Db      459 LeuThrLysLeuThrLysSerGlnSerAspTyrgLysSerIleProIleGlyProTyrg 478
Qy      1438 AGTCTGATGCTTACCGCTTAT 1458
Db      479 LysLeuA-gLeuTyrgLysArgLys 485

```

## RESULT 5

S38379  
 adenosylhomocysteinase (EC 3.3.1.1) - Madagascar periwinkle  
 N/Alternate names: S-adenosyl-L-homocysteine hydrolase  
 C/Species: Catharanthus roseus (Madagascar periwinkle)  
 C/Date: 22-Jan-1994 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999

C/Accession: S38379  
 R/Schroeder, G.; Waitz, A.; Hotze, M.; Schroeder, J.  
 submitted to the EMBL Data Library, October 1993  
 A/Description: cDNA for stress-induced S-adenosyl-L-homocysteine hydrolase from Madagascar  
 A/Reference number: S38379  
 A/Accession: S38379  
 A/Molecule type: DNA  
 A/Residues: 1-485 [SCH]  
 A/Cross-references: EMBL:Z26881; NID:g407411; PIDN:CAA81527.1; PID:g407412  
 C/Genetics:  
 A/Gene: SAHH  
 C/Superfamily: adenosylhomocysteinase  
 C/Keywords: NAD; thioether hydrolase  
 F/263-292/Region: beta-alpha-beta NAD nucleotide-binding fold  
 F/86,120/Active site: Cys #status predicted

Alignment Scores:	2,68e-92	Length: 485
Pred. No.:	1520.00	Matches: 304
Score:	74.18%	Conservative: 58
Percent Similarity:	62.30%	Mismatches: 108
Best Local Similarity:	55.54%	Indels: 18
Query Match:	2	Gaps: 6
DB:		

US-09-759-990A-1 (1-1461) x S38379 (1-485)

```

Qy      19  ACTGGTCTCCATTCGAGTACAGAAATGCCGACATCAACCTCCATGTTCTGGCCGCTAAG 78
Db      8  ThrSerSerGlyArgGluTyrgLysValLysAspMetSerGlnAlaAspPheGlyArgLeu 27
Qy      79  GAATTCACCTTCTCTGAGAGAAATGCCAGGCTTATGCTTCTCGTGAGCGCTATTCC 138
Db      28  GluIleGluLeuAlaGluValGluMetProGlyLeuMetSerCysArgAlaGluPheGly 47
Qy      139  GCTTCTAAGCCATTGAAGGTGTGAGAACTCTGCTGTTCTCCATCAGATCAGAGTCCAGACA 198
Db      48  ProSerGlnProPheLysGlyAlaLysIleThrGlySerLeuHisMetThrIleGlnThr 67
Qy      199  GCCGTCTCATTCAGACACTCATAGCTCTTGGTGTGTGATGTCAGATGGGCTTCTGCAAC 258
Db      68  AlaValLeuIleGluThrLeuThrAlaLeuGlyAlaGluValArgTrpCysSerCysAsn 87
Qy      259  ATCTTCTCTACACAGATACAGCCGCTGCTGCTATCTGTTGTCGGCCCAACAGGACACCA 318
Db      88  IlePheSerThrGlnGlnHisAlaAlaAlaIleAlaArgAspSerAlaAla----- 405
Qy      319  GAGAAGCCAGCCGCTATCCAGTCTTCCGCTGGAAGGCGGAAACACTCCAGATACTGG 378
Db      106  -----ValPheAlaTrpLysGlyGluThrLeuGlnGluTyrgTrp 118
Qy      379  GAGAACACATACCGCGCTCTCATATGG---CCAGATGGTCAAGGCCACAGAGTTGTC 435
Db      119  TrpCysThrGluArgAlaLeuAspTrpGlyProAspGly---GlyProAspLeuIleVal 137
Qy      436  GATCATCGTGGTCACTCATCTCTCATCTCCAGGGGCTTC-----GAATTC 483
Db      138  AspAspGlyGlyAspAlaThrIleLeuIleHisGluGlyValLysAlaGluGluGluTyrg 157
Qy      484  GAAACACCGCGTCTGTTCTCCAGAGCCAAACAGAGCTGACACCTCGAATACCGCTGGTT 543
Db      158  LysLysAsnGlyAlaLeuProAspProSerSerThrAspAsnAlaGluPheGlnIleVal 177
Qy      544  CTTCTCTACTCAAGCAGGTCTTCAACCAAGACAGAACCACTGGCACACAGTTGCTGCC 603
Db      178  LeuThrIleIleArgAspGlyLeuLysSerAspProThrLysTyrgThrArgMetLysGlu 197
Qy      604  GGCATGAACGCTGTTTCCGAGAGACAAACACAGGTGTCCACCGCTCTACAGCTCGAG 663
Db      198  ArgLeuValGlyValSerGluGluThrThrGlyValLysArgLeuTyrgLysMetGln 217
Qy      664  AAGAGGCGCAAACTCTCTTCCAGCCCATCAACGTCACAGCTGTGTACAAAGTCCAG 723
Db      218  AlaAsnGlyThrLeuLeuPheProAlaIleAsnValAsnAspSerValThrLysSerLys 237

```

724 TTCGATACATCTACGGCTCCGCCACTCCCTTATCGATGATACACCGTCTCCGAT 783  
Db PheAspAsnLeuTyrGlyCysArgHisSerLeuProAspGlyLeuMetArgAlaThrAsp 257  
784 GTCATGATCGCGCAGACAGAGCTCTCGTCATGGTTACGGCATCTCGCAAGGCTGC 843  
Db ValMetIleAlaGlyLysValAlaValAlaGlyTyrGlyAspValGlyLysGlyCys 277  
784 GTCATGATCGCGCAGACAGAGCTCTCGTCATGGTTATCATCACAGAACTCGACCAATCTGC 903  
Db AlaAlaAlaLeuLysGlnAlaGlyAlaArgValIleValThrGluIleAspProIleCys 297  
904 GCTCTCCGGCTGCCATGAGGCTACACAGGCTCCCGCATCTCAGAGATCTCAGAT 963  
Db AlaLeuGlnAlaThrMetGluGlyLeuGlnValLeuThrLeuGlnAspValValSerGlu 317  
964 GTCGATATCTCTGTTACATGCACAGAACTCGCATATCATCTCTGTCACATGATGCC 1023  
Db AlaAspIlePheValThrThrGlyAsnLysAspIleIleMetValAspHisMetArg 337  
1024 CAGATGAGGATAGGCTATTGCGGTAACTCGGCCACTTCGATACGAAATTTGATACA 1083  
Db LysMetLysAsnAsnAlaIleValCysAsnIleGlyHisPheAspAsnGluIleAspMet 357  
1084 GATGGCTCTCATGAATACCCAGGATCAGACATCCCATCAGCCAGATACACATG 1143  
Db LeuGlyLeuGluThrThrProGlyValLysArgIleThrLysProGlnThrAspArg 377  
1144 TGGGAATCCGATGCGCCACGCT--ATCCTCTCTTCTGTCAGGCGCGCTCTTAAAC 1200  
Db TrpValPheProAspThrAsnSerGlyIleIleValLeuAlaGluGlyArgLeuMetAsn 397  
1201 CTTGCTCGCTCAGCTACCCATCTTTGTTATGTCATGTCATCAACCAACAGACA 1260  
Db LeuGlyCysAlaThrGlyHisProSerPheValMetSerCysSerPheThrAsnGlnVal 417  
1261 CTCGCTCAGCTCAGCTCTACGAAAGAGA-----GGAAATCTCGAGAGAGGTTTAC 1314  
Db IleAlaGlnLeuGluLeuThrAsnGluArgLysThrGlyLysTyrGluLysValTyr 437  
1315 ACATCTCGAAGCATCTCGATGAAGTCGCTCGCTCCACCTCGGATCTCTCGATGC 1374  
Db ValLeuProLysHisLeuAspGluLysValAlaAlaLeuHisLeuGlyLysLeuGlyAla 457  
1375 CACCTTACAAGCTTACAGAGAGGCTGACATACATCAACCTTCAGTTGAGGTCCT 1434  
Db LysLeuThrLysLeuThrLysAspGlnAlaAspTyrIleSerValProIleGluGlyPro 477  
1435 TACAAGTCTGATGCTTACCGTTAT 1458  
Db TyrLysProAlaHisTyrArgTyr 485  
RESULT 6  
S71621  
adenosylhomocysteine (EC 3.3.1.1) - Phalaenopsis sp.  
N:Alternate names: S-adenosyl-L-homocysteine hydrolase  
C:Species: Phalaenopsis sp.  
C:Date: 27-Nov-1997 #sequence\_revision 12-Dec-1997 #text\_change 22-Jun-1999  
C:Accession: S71621  
R:Preisig-Mueller, R.; Gnaul, P.; Kindl, H.  
Arch. Biochem. Biophys. 317, 201-207, 1995  
A:Title: The inducible 9,10-dihydrophenanthrene pathway: characterization and expression  
A:Reference number: S71619; MUID:95177653; PMID:7872785  
A:Accession: S71621  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-485 <PRE>  
A:Cross-references: EMBL:X79905; NID:g758246; PIDN:CAA56278.1; PID:g758247  
A:Experimental source: clone SH511  
C:Superfamily: adenosylhomocysteinease  
C:Keywords: NAD; thioether hydrolase  
F:263-292/Region: beta-alpha-beta NAD nucleotide-binding fold

F:86,120/Active site: Cys #status predicted

Alignment Scores:  
Pred. No.: 1,42e-91 Length: 485  
Score: 1509.00 Matches: 302  
Percent Similarity: 73.44% Conservative: 52  
Best Local Similarity: 62.66% Mismatches: 112  
Query Match: 55.13% Indels: 16  
DB: 2 Gaps: 4

US-09-759-990A-1 (1-1461) x S71621 (1-485)

QY 34 GAGTACAGATTCGACATCAACCTCATGTTCTGGCGCTAAGAACTTACCCCTGCT 93  
Db 13 GluTyrLysValLysAspLeuSerGlnAlaAspPheGlyArgLeuGluIleGluLeuAla 32  
QY 94 GAGAAAGAAATGCCAGGTCTTATGTTCTTCTGAGCGTTATTCCGCTTCAAGCCATTG 153  
Db 33 GluValGluMetProGlyLeuMetAlaCysArgAlaGluPheGlyProSerGlnProPhe 52  
QY 154 AAGGTGTGCAATCTCTGTTCCCTCCACATGACATCCACAGCCGCTCTCATCGAG 213  
Db 53 LysGlyAlaArgIleSerGlySerLeuHisMetThrIleGlnThrAlaValLeuIleGlu 72  
QY 214 ACATCTACAGCTCTTGGTCTCATGTCAGATGCGCTTCTTCCGACATCTTCTTACACAA 273  
Db 73 ThrLeuThrAlaLeuGlyAlaGluValArgTrpCysSerCysAsnIlePheSerThrGln 92  
QY 274 GATACAGCGCTCTCATGTTCTGTCGCGCCACAGGCACACAGAGAGCCAGCGGT 333  
Db 93 AsphHisAlaAlaAlaIleAlaArgAspSerAlaAla----- 105  
QY 334 ATCCCATGCTTCGCTCGAAGGCGAANAACCTCCAGAACTACTGGGAGAACACATACCG 393  
Db 106 -----ValPheAlaThrLysGlyLysGluThrLeuGlnLysTrpTrpCysThrGluArg 123  
QY 394 GCTCTCATGCGCCAGATGTCAGGCCACAGAGCTTCTCGATGATGCTGCTGCTGCT 453  
Db 124 CysLeuGluTrpGlyAlaGlyGlyProAspLeuIleValAspGlyGlyAspAla 143  
QY 454 ACATCTCTCATCTCCAAAGGCTTC-----GAATTCGAACACCGCGCTGTT 501  
Db 144 ThrLeuLeuIleHisGluGlyValLysAlaGluGluTyrGlyLysAsnGlyLysIle 163  
QY 502 CGAGAGCAACAGAGCTGCAACCTCGAATACGCTGCTTCTTCTTACTACTCAAGCAG 561  
Db 164 ProAspProAlaSerThrAspAsnAlaGluPheGlnIleValLeuGlyLeuIleArgAsp 183  
QY 562 GTCTTCAACCAAGACAGAACCCACTGGGCACACAGTTCCTGCCGCATGAACGCTGTTCC 621  
Db 184 SerLeuSerValAspProLysTyrArgArgMetLysGluArgLeuValGlyValSer 203  
QY 622 GAAGAGACAACACAGGTTCACCGCTTACCGCTTACAGAGAGGAGGCGCAACTCCTC 681  
Db 204 GluGluThrThrGlyValLysArgLeuTyrGlnMetGlnTyrSerGlyThrLeuLeu 223  
QY 682 TTCACGACCTCAAGCTCAACGCTGTTACAAAGTCCAGTTCGATCAATCAATCAAGC 741  
Db 224 PheProAlaIleAsnValAsnAspSerValThrLysSerLysPheAspAsnLeuTyrGly 243  
QY 742 TGGCGCCACTCCCTTATCGATGTTATCAACGCTTCCGATGTCATGTCATCGCGGCAAG 801  
Db 244 CysArgHisSerLeuProAspLysLeuMetArgAlaThrAspValMetIleAlaGlyLys 263  
QY 802 ACAGCTCTCTCATGTTTACGGCATGTCGCAAGGCTCGCTCAATCTCCCTCGGTGC 861  
Db 264 ValAlaValValCysGlyTyrGlyAspValGlyLeuGlyCysAlaAlaLeuLysThr 283  
QY 862 CAGGCGCTCGCTTATCATACAGACTCGACCCCAATCTCGCTCTCTCCAGCTGCGCATG 921  
Db 284 AlaGlyAlaArgValIleValThrGluIleAspProIleCysAlaLeuGlnAlaLeuMet 303  
QY 922 GAAGGCTACCGAGTCCGCGCATCGAGGAGTCTCAAGAGTGTGATATCTTCGTACA 981

```
Db 304 GluGlyLeuProValLeuArgLeuGluAspValValSerGluAlaAspIlePheValThr 323
Qy 982 TCACAGGAAGAACTCGGATATCATCTCTGTTGACATGATGCCCGACAGTGAAGTAAGGCT 1041
Db 324 ThrThrGlyAsnLysAspIleLeuMetValAspHisMetArgLysMetLysAsnAspAla 343
Qy 1042 ATTGCGGTAACTACGCGCACTTCGTATAACGAAATTCATACAGATGGCCCTCATCAATAC 1101
Db 344 IleValCysAsnIleGlyHisPheAspAsnGluIleAspMetLeuGlyLeuSerPhe 363
Qy 1102 CCAGCATCAGCAGCATCCCAATCAATCAGCAGCAATACAGATGGGAATTCAGATGGC 1161
Db 364 ProGlyValLysArgIleThrIleLysProGlnThrAsnGlnValPheProAspThr 383
Qy 1162 CACGCT--ATCCCTCTTCTGCTGAGGCGCGCTTCTTAACCTTGGCGCTACAGGT 1218
Db 384 AsnSerGlyIleLeuValLeuAlaGluGlyArgLeuMetAsnLeuGlyCysAlaThrGly 403
Qy 1219 CACCATCTTTCGTTATGTCATGTCATTCACAAACAGACATCTGCTCAGCTCAGCTC 1278
Db 404 HisProSerPheValMetSerSerPheThrAsnGlnValIleAlaGlnLeuGluLeu 423
Qy 1279 TAGCAAAAGAGA-----GGAAATCTCGAGAAGAAGTTTACACACTTCCGAAGCATCTC 1332
Db 424 TrpLysGluArgAlaSerGlyLysTyrgluLysLysValTyrgluLeuProLysHisLeu 443
Qy 1333 GATGAAGAAAGTCGCTCCCTCCACCTCGATCTCTCGATCTCCACCTTACAAAGCTTACA 1392
Db 444 AspGluLysValAlaAlaLeuHisLeuGlyLysLeuGlyAlaLysLeuThrLysLeuThr 463
Qy 1393 CAGAGAGCGGCTGACTACATCAACGTTCCAGTTGAGGCTCTTACAGTCTGATGCTTAC 1452
Db 464 ProSerGlnAlaAspTyrlleSerValProValGluGlyProTyrlleProAlaHisTy 483
Qy 1453 CGTTAT 1458
Db 484 ArgTyr 485

RESULT 7
adenosylhomocysteinase XF1037 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: D82730
C:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; PMID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: D82730
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-446 <SIM>
A:Cross-references: GB:AB003941; GB:AE003849; NID:59105978; PIDN:AAF8347.1; GSPDB:GN001
B:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Reto, E.; Docena, C.; El-Dorry, H.; Facincan, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.B.; Laig
chado, M.A.; Madeira, M.A.B.N.; Vadeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, B.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, P.R.; da Silva, F.R.; da Silva, W.A.; da Silv
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1037
C:Superfamily: adenosylhomocysteinase
```

Alignment Scores:  
Pred. No.: 8 45e-85 Length: 446  
Score: 1406.00 Matches: 277  
Percent Similarity: 73.74% Conservative: 60  
Best Local Similarity: 60.61% Mismatches: 104  
Query Match: 51.37% Indels: 16  
DB: 2 Gaps: 5

US-09-759-990A-1 (1-1461) x D82730 (1-446)

```
Qy 103 ATGCAGGCTCTATGTTCTTCGTCGAGCGTTATTCGCTTCTTAAGCCATTCGAGGCTGTC 162
Db 1 MetProGlyLeuMetSerIleArgArgLysAlaSerLysGlnProLeuLysGlyVal 20
Qy 163 AGAATCTCTGTTCCCTCCACATGACAGTCCAGACGCGTCTCTCATCGACACATCACA 222
Db 21 ArgValThrGlySerLeuHisMetThrIleGlnAlaValLeuIleGluThrLeuLys 40
Qy 223 GCTCTGCTCTGATGTCAGATGGGCTTCTGCAACATCTTCTTACACAGATACAGCC 282
Db 41 AspIleGlyAlaAspValArgTrpAlaSerCysAsnIlePheSerThrGlnAspHisAla 60
Qy 283 GCTGCTCTATCTGTTCTCGGCCCAACAGGCGACACACAGACAGCCGCGTATCCCGTTC 342
Db 61 AlaAlaAlaIleAlaThr-----SerglyThrProVal 71
Qy 343 TTCCTCTGGAAGGCGCAAACTCCACAGATACTGGGAGAACACATACCGCGCTCTCACA 402
Db 72 PheAlaTrpLysGlyGluThrLeuGluGluTyrrTrpAspCysThrLeuGlnAlaLeuThr 91
Qy 403 TGG-----CCAGATGGT-----CAAGGCCACACAGAGTTGTTCGATGATGTTGGTAT 450
Db 92 PheThrLeuSerAspGlyThrLeuThrGlyProGluLeuIleValAspAspGlyGlyAsp 111
Qy 451 GCTACACTCTCTCATCTCAAGGCTTCGAATTCGAAACAGACGCGGTGCT---GTTCCAGAG 507
Db 112 AlaThrLeuLeuIleHisLysGlyTyrgluLeuGluAsnGlySerThrTrpValAspGlu 131
Qy 508 CCAACAGACGCTGACAACTCGAATACCGCTGGTTCCTTTCCTACACTCAAGCAGGTTCTC 567
Db 132 Pro-----SerAspSerLeuGluGluGlnValIleLysAtgLeuLeuLysAtgIleAla 149
Qy 568 AACCAACAGACAGAACCACTGGCGACACAGTTCGCGGCATCAGCGTGTTCGAGAGAG 627
Db 150 IleGluArgProGlyTyrrTrpThrArgValValAsnAspTrpLysGlyValSerGluGlu 169
Qy 628 ACAACAAACAGGTCTCCACCGCTCTACAGCTCGAAGAGGCGCAAACTCTTCTCCCA 687
Db 170 ThrThrThrGlyValHisArgLeuTyrgluIleAlaAlaThrGlyArgLeuLeuValPro 189
Qy 688 GCGATCAACTCAACAGCGCTGTTACAAAGTCCAAAGTTCGATTAACATCTACGCTCCGCG 747
Db 190 AlaIleAsnValAsnAspSerValThrLysSerLysPheAspAsnLeuTyrglyCysArg 209
Qy 748 CACTCCCTCTATCGATGTTCAACCGCTTCGATGTCATCATCGCGCGCAGACAGCT 807
Db 210 GluSerLeuAlaAspGlyLeuLysArgAlaMetAspValMetLeuAlaGlyLysLeuAla 229
Qy 808 CTCCTCATGTTAGCGGCTGTCGGAAGGCTCGGCTCAATCCCTCCGTGGGCCAAGGC 867
Db 230 ValValCysGlyTyrglyAspValGlyLysGlySerAlaHisSerLeuArgAlaTyrgly 249
Qy 868 GCTGCGCTTATCATCAGACTCGACCCCAATTCGCTCTCCAGGTCGCGCATGGAAGGC 927
Db 250 AlaArgValIleValThrGluIleAspProIleCysAlaLeuGlnAlaAlaMetGluGly 269
Qy 928 TACAGAGTCGCGCGATCGAGAGTCTCGTCAAGATGTCGATCTCTCGTTACATCCACA 987
Db 270 PheGluValThrThrValGluAspThrLeuGlyGlnAlaAspIleTyrrValThrThr 289
Qy 988 GGAAACTCGCATCATCTCTGTTGACATGATGCGCCAGATGAAGGATGAAGGCTATTGTC 1047
```

Db 290 GlyAsnLysAspValIleAArgIleGluHisMetThrAlaMetLysAspGlnValIleVal 309  
 Qy 1048 GGTATCATCGGCCTCTCGATACGAAATGATACAGATGGCTCTATGAATACCCAGGC 1107  
 Db 310 CysAsnIleGlyHisPheAspAsnGluIleGlnValAspAlaLeuAsnAlaLeuAlaGly 329  
 Qy 1108 ATCAAGCACATCCCAATCAAGCCAGAAATACGACATGTGGGAATTCACAGATGCCACGCT 1167  
 Db 330 ValGlnLysIleAsnIleLysProGlnValAspLysPheIleLeuProAsnGlyAsnThr 349  
 Qy 1168 ATCTCTCTTGTGTAGGCGCCCTCTTAACCTTGGCTGCGCTACAGGTCAACCATCT 1227  
 Db 350 LeuPheLeuLeuAlaGluGlyArgLeuValAsnLeuGlyCysAlaThrGlyHisProSer 369  
 Qy 1228 TTCGTATGTCATTCATTCACAAACGACGACATCGCTGAGCTCGACCTCTACGAAAG 1287  
 Db 370 PheValMetSerAsnSerPheAlaAsnGlnThrLeuAlaGlnIleAspLeuTrpGlnAsn 389  
 Qy 1288 AGAGGAATCTCGAGAAGAAGTTTACACATTCGGAAGCATCTCGATGAAGAGTGGCT 1347  
 Db 390 LysAspValTyrgluLysAsnValTyrgluLeuProLysLysLeuAspGluValAla 409  
 Qy 1348 CCCTCCACCTCGGATCTCTCGATGCCACCTTACAAAGCTTACACAGAAGCAGCTGAC 1407  
 Db 410 ArgLeuHisLeuGluLysIleGlyValLysLeuThrThrLeuThrAlaAsnGlnAla 429  
 Qy 1408 TATCATCAGTTCAGTGTAGGCTCTTACAGTCTGATGCTTACGTTAT 1458  
 Db 430 TyrLeuGlyIleSerValGluGlyProPheLysProGluHisTyrglyArgTyr 446

## RESULT 8

AF2580  
 S-adenosylhomocysteine hydrolase ahcy [imported] - Agrobacterium tumefaciens (strain C58  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 C:Accession: AF2580  
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayin, T.; Levy, R.; Li, M.; McClell  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
 ster, B.W.  
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; MUID:21608550; PMID:11743193  
 A:Accession: AF2580  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-466 <KUR>  
 A:Cross-references: GB:AB008688; PIDN:AAL41060.1; PID:g17738347; GSPDB:GN00186  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: ahcy  
 A:Map position: circular chromosome  
 C:Superfamily: adenosylhomocysteinase

## Alignment Scores:

Pred. No.: 3,25e-83 Length: 466  
 Score: 1382.00 Matches: 280  
 Percent Similarity: 71.13% Conservative: 60  
 Best Local Similarity: 58.58% Mismatches: 118  
 Query Match: 50.49% Indels: 20  
 DB: 2 Gaps: 6

US-09-759-990A-1 (1-1461) x AF2580 (1-466)

Qy 34 GAGTACAGAAATCCGACATCAACCTCCATGTTCTCGCGCGTAAGGAATTCACCTTGCT 93  
 Db 6 AspTyrlleValAlaAspIleAsnLeuAlaAlaPheGlyArgLysGluLeuAspIleAla 25  
 Qy 94 GAGAGGAAATCCAGGTCTTATGTTCTGTCGAGGTATTCGCGCTTCAAGCCATTG 153  
 Db 26 GluThrGluMetProGlyLeuMetSerCysArgLysGluPheGlyGluSerLysProLeu 45

Qy 154 AAGCGTGCAGATCTCTGGTTCCCTCCACATGACAGTCCAGACAGCCGCTCTCATCGAG 213  
 Db 46 LysGlyAlaArgIleThrGlySerLeuHisMetThrIleGlnThrGlyValLeuIleGlu 65  
 Qy 214 ACATCTCAGCTCTTGGTGTCTGATGTCAGATGGCTTCTCGAAACATCTTCTTACACAA 273  
 Db 66 ThrLeuLysGluLeuGlyAlaGluIleArgTrpAlaSerCysAsnIlePheSerThrGln 85  
 Qy 274 GATCAGCGCTGCTGCTATTCGTTCTCGGCCCAACAGGCACACACAGAGAACGCCGCGT 333  
 Db 86 AspHisAlaAlaAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIle 96  
 Qy 334 ATCCAGCTCTCCCTCGAGGCGGAACACTCCACGATACTGGGAGNACACATACCGC 393  
 Db 97 IleProValPheAlaValLysGlyGluSerLeuThrGluTyrrpGluTyrrpAspLys 116  
 Qy 394 GCTCTCATGTCGACATGTCAGGCCACACGACGCTGTCGATGATGGTGGTGATGCT 453  
 Db 117 IlePheGlnTrpThrAspGlyGlyLeuSerAsnMetIleLeuAspAspGlyGlyAspAla 136  
 Qy 454 ACATCTCTCATCTCCAGGCTTCGAATTCGAACAGCC---GGTGTGTTCACAGGCCA 510  
 Db 137 ThrMetTyrlleLeuLeuGlyAlaArgAlaGluAlaGlyLysValLeuSerAsnPro 156  
 Qy 511 ---ACAGAAGCTGACAACTCGAATACCGCTGCGTTCCTTCTACACTCAAGCAGCTCTC 567  
 Db 157 GlySerGluGluGluGluIleLeuPheAlaGlnIleAsnLysArgLeuLysAla----- 174  
 Qy 568 AACCAAGACAGAACCACTGGCACACAGTCTGCTGCC---GGCATGAACGGTGTTCGGAA 624  
 Db 175 -----SerProGlyTrpPheThrLysGlnArgAspAlaLeuLysGlyValThrGlu 191  
 Qy 625 GAGACACACAGGTGTCCACGCTCTACAGTTCGAGAGAGGGCAACTCTCTTTC 684  
 Db 192 GluThrThrThrGlyValHisArgLeuTyrrAspLeuAlaLysLysGlyLeuLeuProPhe 211  
 Qy 685 CCAGCCATCAAGTCAACACGCTCTTAAAGTCCAAAGTTCGATAACATCTACGGTGC 744  
 Db 212 ProAlaIleAsnValAsnAspSerValThrLysSerLysPheAspAsnLysTyrglyCys 231  
 Qy 745 CGCCACTCCCTTATCGATGCTTCAACGCTCTTCCGATGTCATGATCGCGGCAAGACA 804  
 Db 232 LysGluSerLeuValAspGlyIleArgAlaThrAspValMetMetAlaGlyLysVal 251  
 Qy 805 GCTCTGCTCATGGGTACCGCATCTCGGCAAGGCTCGCTCAATCCCTCGTGGCCAA 864  
 Db 252 AlaValValCysGlyTyrglyAspValGlyLysGlySerAlaAlaSerLeuGlnGlyAla 271  
 Qy 865 GCGCTGCGTTCATCATCAGACTCGACCAATCTCGCTCTCCAGCTGCCATGAA 924  
 Db 272 GlyAlaArgValLysValThrGluIleAspProIleCysAlaLeuGlnAlaAlaMetAsp 291  
 Qy 925 GGCTACAGGTTCGCGCATCATCTCTGTTACATGATGCGCCAGATGAGGATTAAGCTATT 1044  
 Db 292 GlyPheGluValValArgLeuGluAspValValSerSerAlaAspIlePheIleThrThr 311  
 Qy 985 ACAGGAACCTCCGATCATCTCTGTTACATGATGCGCCAGATGAGGATTAAGCTATT 1044  
 Db 312 ThrGlyAsnLysAspValIleArgIleGluHisMetArgGluMetLysAspMetAlaIle 331  
 Qy 1045 GTCGGTAACATCGGCCACTTCGATAACGAAATTCATACATGCTCATGAATATCCCA 1104  
 Db 332 ValGlyAsnIleGlyHisPheAspAsnGluIleGlnValAlaSerLeuArg----- 348  
 Qy 1105 GGCAATCAAGCACATCCCAATCAAGCCAGATGACATGTTGGGAATTCACAGATGGCCAC 1164  
 Db 349 AsnLeuLysTrpThrAsnIleLysProGlnValAspMetIleGluPheProLysGlyAsn 368  
 Qy 1165 GCTATCTCTCTTCTGCTCAGGGCGCCCTTCTTAACCTGGCTGGCTACAGGTCAACCA 1224  
 Db 369 ArgIleIleLeuLeuSerGluGlyArgLeuLeuAsnLeuGlyAsnAlaThrGlyHisPro 388  
 Qy 1225 TCTTTGCTGATGTCAATGTCTATTCAAAACACAGACACTCGCTCAGCTCGACTCTACGAA 1284

```

Db      389 SerPheValMetSerAlaSerPheThrAenGlnValLeuGlyGlnIleGluLeuPheThr 408
Qy      1285 AAGAGAGAAATCTCGAAGAAGAGTTTACACATCTCCGAAGCATCTCGATGAAGAAGTC 1344
Db      409 LysProGlyGluTyrLysAenGluValTyrValLeuProLysHisLeuAspGluLysVal 428
Qy      1345 GTCGCGCTCCACCTCGATCTTCGATGCCACCTTACAAGCTTACAGAGTTCACAGAGCAGCT 1404
Db      429 AlaArgLeuHisLeuGluLysLeuGlyValArgLeuThrGluLeuSerAspLeuGlnAla 448
Qy      1405 GACTCATCAACAGTCCAGTTGAGGCTCTTACAAAGTCTGATGCTTACCGTTAT 1458
Db      449 AspTyrIleGlyIleSerLysGlnGlyProPheLysAlaGluHisTyrArgTyr 466

RESULT 9
D97362
adenosylhomocysteine (S-adenosyl-L-homocysteine hydrolase) [adobcyase] - Ag
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C:Accession: D97362
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: D97362
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-466 <KUR>
A:Cross-references: GB:AB007869; PIDN:AAK85853.1; PID:g15154892; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_46
A:Map position: circular chromosome
C:Superfamily: adenosylhomocysteine

Alignment Scores:
Pred. No.:      3 25e-83      Length:      466
Score:          1382.00      Matches:     280
Percent Similarity: 71.13%      Conservative: 60
Best Local Similarity: 58.58%      Mismatches: 118
Query Watch:      50.49%      Indels:      20
DB:              2          Gaps:         6

US-09-759-990A-1 (1-1461) x D97362 (1-466)

Qy      34  GAGTACAGATTCGCGCATCACTCCATGTTCTCGCGCGTAAGAACTTACCTTGCT 93
Db      6  AspTyrIleValAlaAspIleAenLeuAlaAlaPheGlyArgLysGluLeuAspIleAla 25
Qy      94  GAGAAGGAATCCAGAGCTTATGTTCTTCGTGAGCGTTATTCGCTTCAAGCCATTG 153
Db      26  GluThrGluMetProGlyLeuMetSerCysArgLysGluPheGlyGluSerLysProLeu 45
Qy      154  AAGGGTGTCAAGATCTCTGTTCCCTCCATGACAGTCCAGACAGCGCTCTCATCGAG 213
Db      46  LysGlyAlaArgIleThrGlySerLeuHisMetThrIleGlnThrGlyValLeuIleGlu 65
Qy      214  AACTCAGAGCTTGTGTGTGATGATGATGAGTGGGCTTCCTGCAACATCTCTTACAA 273
Db      66  ThrLeuLysGluLeuGlyAlaGluIleArgTrpAlaSerCysAsnIlePheSerThrGln 85
Qy      274  GATACAGCGCTGTGTATCGTTGTGTCGCCCAACAGGCACACCAGAGAAGCCGCGT 333
Db      86  AspHisAlaAlaAlaAlaIleAlaAla-----AlaGly 96
Qy      334  ATCCAGATCTCGCTCGAAGGGCGAAACACTCCCAAGTACTGGGAGAACACATACCGC 393
Db      97  IleProValPheAlaValLysGlyGluSerLeuThrGluTyrTrpGluTyrThrAspLys 116
Qy      394  GCTCTCAGATGCCAGATGTTCAAGGCCACAGCAGGTGTTCGATGATGCTGATGCT 453
Db      117  IlePheGlnTrpThrAspGlyGlyLeuSerAsnMetIleLeuAspGlyGlyAspAla 136

```

RESULT 10

H87280

adenosylhomocysteine [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

```

Qy      454  ACATCTCTCATCTCCAGAGGCTTCGAATTCGAACAGCC---GGTCTGTTCAGAGCCA 510
Db      137  ThrMetTyrIleLeuLeuGlyAlaArgAlaGluValLeuAspValLeuSerAsnPro 156
Qy      511  ---ACAGAGCTGACAACTCGAATACCGCTGCTTCTTCTACACTCAAGAGGTCTTC 567
Db      157  GlySerGluGluGluIleLeuPheAlaGlnIleAsnLysArgLeuLysAla----- 174
Qy      568  AACCAAGACAGAACCACTGGCACACAGTTGCTGCC---GGCATGAACGGTGTTCGAA 624
Db      175  -----SerProGlyTrpPheThrLysGlnArgAspAlaLeuLysGlyValThrGlu 191
Qy      625  GAGACACACAGGTGTCCACGCTCTTACCATGTCGAGAGGAGGCAAACTCTCTTC 684
Db      192  GluThrThrThrGlyValHisArgLeuTyrAspLeuTyrAspLeuAlaLysLysLeuProPhe 211
Qy      685  CCAGCCATCAACGTCAACGACCTGTACAAGTTCGAAGTTCGATAACATCTACGGTGC 744
Db      212  ProAlaIleAsnValAsnAspSerValThrLysSerLysPheAsnLysTyrGlyCys 231
Qy      745  CGCCACTCTCTTATCGATGTATCAACCGTCTTCCGATGTCTCATGTCGCGGCAAGACA 804
Db      232  LysGluSerLeuValAspGlyIleArgAlaThrAspValMetMetAlaGlyLysVal 251
Qy      805  GCTCTCGTCATGGGTACCGCATGTCCGCGAGGCTCGCTCAATCCCTCGTGGCCAA 864
Db      252  AlaValValCysGlyTyrGlyAspValGlyLysGlySerAlaAlaSerLeuGlnGlyAla 271
Qy      865  GCGCTCGCTTATCATCAAGAACTCCAGCACTTCGCTCTCCAGCTGCCATGAA 924
Db      272  GlyAlaArgValLysValThrGluLeaspProIleCysAlaLeuGlnAlaAlaMetAsp 291
Qy      925  GCTTACAGGTTCGCGCATCGAGGAAGTCTCAAGANGTCGATATCTTCTTACATGC 984
Db      292  GlyPheGluValValArgLeuGluAspValValSerSerAlaAspIlePheIleThr 311
Qy      985  ACAGGAACCTCGATATCATCTCTGTTCACATGATGCCGCCAGATGAAGTAAGCTATT 1044
Db      312  ThrGlyAsnLysAspValIleArgIleHisMetArgLysMetLysAspMetAlaIle 331
Qy      1045  GTCGGTAACATCGGCCACTTCGATAACGAATTCATACAGATGGCTCATGAATACCCA 1104
Db      332  ValGlyAsnIleGlyHisPheAspAsnGluIleGlnValAlaSerLeuArg----- 348
Qy      1105  GCATCAAGCACATCCCAATCAAGCCAGATAACATGTGGGAATTCACAGATGGCAC 1164
Db      349  AsnLeuLysTrpThrAsnIleLysProGlnValAspMetIleGluPheProLysGlyAsn 368
Qy      1165  GCTATCTCTCTTCTGCTGAGGCGCGCTTCTTAACCTTGGCTGCGCTACAGTCAACCA 1224
Db      369  ArgIleIleLeuLeuSerGluGlyArgLeuLeuAsnLeuGlyAsnAlaThrGlyHisPro 388
Qy      1225  TCTTCTGTTATGTCAATGTCAATTCACAAACACAGACTCTGCTCAGCTCGACTTACGAA 1284
Db      389  SerPheValMetSerAlaSerPheThrAsnGlnValLeuGlyGlnIleGluLeuPheThr 408
Qy      1285  AAGAGAGAAATTCGAGAGAAGGTTTACACTCCGAGAGCATCTCGATGAAGAAGTC 1344
Db      409  LysProGlyGluTyrLysAsnGluValTyrValLeuProLysHisLeuAspGluLysVal 428
Qy      1345  GCTGCGCTCCACCTCGGATCTCTCGATGTCCTCAACCTTCAAAAGCTTACACAGACAGCT 1404
Db      429  AlaArgLeuHisLeuGluLysLeuGlyValArgLeuThrGluLeuSerAspLeuGlnAla 448
Qy      1405  GACTACATCAACGTTCACGTTGAGGTCCTTACAGTCTGATGCTTACCGTTAT 1458
Db      449  AspTyrIleGlyIleSerLysGlnGlyProPheLysAlaGluHisTyrArgTyr 466

RESULT 10
H87280
adenosylhomocysteine [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus

```

Cdate: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C/Accession: H87280  
R/Nierman, W.C.; Felblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
N. J.; Ermolaeva, K.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A/Title: Complete Genome Sequence of *Caulobacter crescentus*.  
A/Reference number: A87249; MUID:21173698; PMID:11259647  
A/Accession: H87280  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-463 <STO>  
A/Cross-references: GB:AE005673; MID:g13421390; PIDN:AAK2244.1; GSPDB:GN00148  
C/Genetics:  
C/Superfamily: adenosylhomocysteinase

Alignment Scores:  
Pred. No.: 1.09e-82 Length: 463  
Score: 1374.00 Matches: 276  
Percent Similarity: 70.95% Conservative: 61  
Best Local Similarity: 58.11% Mismatches: 124  
Query Match: 50.20% Indels: 14  
DB: 2 Gaps: 3

US-09-759-990A-1 (1-1461) x H87280 (1-463)

```
QY 34 GAGTACAGAAATCCGACATCAACCTCCATGTTCTCCGCGTGAAGAACTTACCCCTGCT 93
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 3 AspTyrIleValLysAspIleSerLeuAlaAspPheGlyArgLysGluIleAlaIleAla 22

QY 94 GAGAAGAAATCCAGGCTTATGGTCTTCGTGAGCGTTATCCGCTTCAAGCCATG 153
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 23 GluThrGluMetProGlyLeuMetAlaThrArgAlaGluTyrGlyProGlnGlnIleLeu 42

QY 154 AAGGGTGTGAGATCTCTGGTTCCTCCATGACAGTCCAGACAGCCGCTCTCATCCAG 213
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 43 LysGlyAlaArgIleAlaGlySerLeuHisMetThrIleGlnThrAlaValLeuIleGlu 62

QY 214 ACCTCAGAGCTTGTGCTGATGTCAGATGGGCTTCTGCAACATCTTCTCTACAA 273
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 63 ThrLeuThrAlaLeuGlyAlaGluValArgTyrAlaSerCysAsnIlePheSerThrGln 82

QY 274 GATACAGCGCTGCTGCTATCGTTGTCGGCCCAACAGGCACACAGAGAGCCAGCGT 333
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 83 AspHisAlaAlaAlaAlaIleAlaAla-----AlaGly 93

QY 334 ATCCAGTCTTCGCTGGAAGGCGAAACACTCCAGAAATACTGGGAGACACATACCGC 393
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 94 ValProValPheAlaPheLysGlyGluAsnLeuValGluTyrTrpGluTyrAlaHisLys 113

QY 394 GCTCTCATATGCGAGATGTCAGGCCACAGAGTTCGATGATGTTGGTATGCT 453
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 114 IlePheGluThrHisAspGlyGlyTyrProAsnLeuIleLeuAspAspGlyGlyAspAla 133

QY 454 ACATCTCTCATCCAGGCTTCGAATTCGAAACAGCGGCTGTGTCAGAGCAACA 513
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 134 ThrLeuLeuCysValLeuGlyProLysAlaGluLysAspProSerIle-----LeuAsn 151

QY 514 GAAGCTGACAACTCGATACCGCTGGTTCCTGCTTACACTCAACAGAGCTTCACAA 573
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 152 AsnProGlnAsnGluGluGluAlaLeuTyrAlaValMetLysLysTyrLeuAlaGlu 171

QY 574 GACAGAACCACTGGCACACAGTTCGCGCGATCAACGGTTCCTCCGAGAGACACA 633
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 172 LysProGlyPheTyrSerAlaIleArgAlaAlaIleGlyValSerGluGluThrThr 191

QY 634 ACAGGTTCACCGCTCTACAGCTCAGAGAGGGCAACCTCTTCCAGAGCCATC 693
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 192 ThrGlyValHisArgLeuTyrGlnMetAlaGlnLysAspGluLeuProPheProAlaIle 211

QY 694 AACGTCAAGAGCTGTTACAAAGTCCAAAGTTCGATTAACATTCAGCTCCGCCACTCC 753
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
```

```
Db 212 AsnValAsnAspSerValThrLysSerLysPheAspAsnLeuTyrGlyCysArgGluSer 231
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 754 CTTATCGATCGTATCAACCGTCTTCCGATGTCATGATCGCGGCAACAGAGCTCTCGTC 813
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 232 LeuValAspAlaIleArgArgGlyThrAspValMetLeuSerGlyLysValAlaValVal 251
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 814 ATGGGTTACCGCATGTCGGCAAGGCTGGCTCAATCTCCGTGGCCCAAGGCGCTCCG 873
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 252 CysGlyTyrGlyAspValGlyLysGlySerAlaAlaSerLeuArgGlnGlyAlaArg 271
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 874 GTTATCATCATCAGACTCGACCCATCTGGCTCTCCAGCTGCCATGGAAGGCTACACAG 933
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 272 ValIleValThrGluValAspProIleCysAlaLeuGlnAlaMetGluGlyTyrGlu 291
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 934 GTCCGCCCATCGAGAGTCTGTCAGAGATGTCGATATCTTCGTTACATGCACAGGAAC 993
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 292 ValGlnThrLeuAsnAspValAlaAspLysAlaAspIlePheValThrAlaThrGlyAsn 311
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 994 TCGGATATCATCTCTGTCGATGATGCGCCAGATGAGGATAAGGCTATGTCGGTAAC 1053
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 312 LysAspValIleThrValAspAspMetArgLysMetLysAsnAlaIleValCysAsn 331
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 1054 ATCGGCCACTTCGATAACGAAATTCATACAGATGGCTCATGAATAACCCAGGCATCAAG 1113
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 332 IleGlyHisPheAspSerGluIleAlaGlyLeuArgAsnPhe-----Lys 348
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 1114 CACATCCCATCAAGCCAGATACGATGTCGGAATTCAGATGCGCCAGCGCTATCCCTC 1173
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 349 TrpAspGluIleLysProGlnValHisHisValGluPheProAspGlyLysLeuLeu 368
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 1174 CTTCTGTCGAGCGCGCTCTTAACTGCTGGCTGCTACAGGTCACCCATCTTTGCTT 1233
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 369 ValLeuSerGluGlyArgLeuValAsnLeuGlyAsnAlaThrGlyHisProSerPheVal 388
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 1234 ATGTCATGTCTATCAACCAACAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1293
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 389 MetSerAlaSerPheThrAsnGlnThrLeuAlaGlnIleGluLeuTyrThrAsnLysAla 408
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 1294 ATCTCGAGAGAGAGGTTTACACCTCCGAGAGCTCGATGAGAGAGAGAGAGAGAGAG 1353
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 409 LysTyrGluAsnGlnValTyrThrLeuProLysHisLeuAspGluLysValAlaPheLeu 428
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 1354 CACCTCGGATCTCTCGATGCTCCACCTTACAAAGCTTACACAGAGAGAGAGAGAGAG 1413
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 429 HisLeuGluLysLeuGlyAlaLysLeuThrThrLeuArgLysAspGlnAlaAspTyrIle 448
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 1414 AACGTTCCAGTGGAGGCTCTTACAGTCTGATGCTTACCGTTAT 1458
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 449 GlyValProGluAlaGlyProPheLysProAspHisTyrArgTyr 463
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
```

## RESULT 11

A46035  
adenosylhomocysteinase (EC 3.3.1.1) - Rhodobacter capsulatus  
N/Alternate names: S-adenosyl-L-homocysteine hydrolase  
C/Species: Rhodobacter capsulatus  
C/Date: 04-Mar-1994 #sequence\_revision 04-Oct-1996 #text\_change 18-Jun-1999  
C/Accession: A46035; A36863  
R/Sganga, M.W.; Aksent, R.R.; Cantoni, G.L.; Bauer, C.E.  
Proc. Natl. Acad. Sci. U.S.A. 89, 6328-6332, 1992  
A/Title: Mutational and nucleotide sequence analysis of S-adenosyl-L-homocysteine hydro  
A/Reference number: A46035; MUID:92335291; PMID:1631127  
A/Accession: A46035  
A/Molecule type: DNA  
A/Residues: 1-462 <SG>  
A/Note: sequence extracted from NCBI backbone (NCBI:108691, NCBI:108696)  
R/Bugg, J.J.; Sganga, M.W.; Bauer, C.E.  
J. Bacteriol. 176, 61-69, 1994  
A/Title: Nucleotide sequence and characterization of the Rhodobacter capsulatus hvrB ge  
A/Reference number: A36863; MUID:94110241; PMID:8282711  
A/Accession: A36863  
A/Molecule type: DNA  
A/Residues: 1-13 <BUG>  
A/Cross-references: GB:L23836; NID:9577613; PIDN:AAAS3540.1; PID:g577614

## C;Genetics:

A;Gene: ahcy  
 C;Function:  
 A;Description: catalyzes the reversible hydrolysis of S-adenosyl-L-homocysteine to adenosyl-L-homocysteine and L-methionine  
 C;Superfamily: adenosylhomocysteinase  
 C;Keywords: NAD; thioether hydrolase  
 F;247-276/Region: beta-alpha-beta NAD nucleotide-binding fold

## Alignment Scores:

Pred. No.: 2,84e-81 Length: 462  
 Score: 1352.50 Matches: 287  
 Percent Similarity: 69.01% Conservative: 47  
 Best Local Similarity: 59.30% Mismatches: 117  
 Query Match: 49.42% Indels: 33  
 DB: 1 Gaps: 9

US-09-759-990A-1 (1-1461) x A46035 (1-462)

```

QY 34 GAGTACAGAAATGCGGACATCACTCCATGTTCTCGCGCGTAAGAACTTACCCCTGCT 93
Db 3 AspTyrIleValLysAspIleLysLeuAlaGluPheGlyArgLysGluLeuAspIleAla 22
QY 94 GAGAAAGAAATGCGGACATCACTCCATGTTCTCGCGCGTAAGAACTTACCCCTGCT 153
Db 23 GluThrGluMetProGlyLeuMetAlaCysArgGluGluPheGlyProSerGlnProLeu 42
QY 154 AAGGTGTCAAGATCTGTTCTGTTGATGTCAGATGGGTCCTGCAACATCTTCTACAA 213
Db 43 LysGlyAlaArgIleAlaGlySerLeuHisMetThrIleGlnThrAlaValLeuIleGlu 62
QY 214 ACATTCACAGCTCTGTTGTTGATGTCAGATGGGTCCTGCAACATCTTCTACAA 273
Db 63 ThrLeuLysAlaLeuGlyAlaAspValArgIlePheSerCysAsnIlePheSerThrGln 82
QY 274 GATACAGCGCTGCTGTTGATGTCAGATGGGTCCTGCAACATCTTCTACAA 333
Db 83 AspHisAlaAlaAlaIleAlaAlaGly-----Gly 93
QY 334 ATCCAGTCTTCGCTGGAGGCGAAACATCCCGAAGTCTGGAGAGACATACCGC 393
Db 94 ThrProValPheAlaValLysGlyGluThrLeuGluGluTyrTrpAlaTyrThrAspLys 113
QY 394 GCTCTCATCATGGCAGATGTCAGGCGCCACAGCAGGTGTGCGATGATGGTGATGCT 453
Db 114 IlePheGlnPheProGluGlyThr-----CysAsnMetIleLeuAspAspGlyIleAla 132
QY 454 ACATCTCTATCTCCAGGCTTCGAATTCGAACA-----GCCGCTCTGTTCCAGAG 507
Db 133 ThrLeuTyrIleLeuLeuGlyAlaArgValGluAlaGlyGluThrAspLeuIleAlaThr 152
QY 508 CCAACAGAGCTGCACACCTCGAATACCGCTGCTTCTGCTACACTCAGCAGGTCTTC 567
Db 153 ProThrSerGluAspGluVal-----Cys-----LeuPhe 162
QY 568 AACCAAGACAAGAACCCAC-----TGGCACACA---GTTCTGCCCGC 606
Db 163 AsnGlnIleLysLysArgMetValGluSerProGlyTrpPheThrGlnGlnArgAlaAla 182
QY 607 ATGACCGTGTTCGAGAGACACACAGAGTGTCCACCGCTCTACACGCTCGAGAG 666
Db 183 IleLysGlyValSerGluGluThrThrThrGlyValHisArgLeuTyrAspLeuHisLys 202
QY 667 GAGGCAACCTCTTCCAGCGCATCAACGCTGTCACAGCTGTTCACAGTCCAGTTC 726
Db 203 LysGlyLeuLeuProPheProAlaIleAsnValAsnAspSerValThrLysSerLysPhe 222
QY 727 GATAACATCATCGCTGCCCGCACTCCCTTATCATGATGATCAACCGCTCTCCGATCTC 786
Db 223 AspAsnLysTyrGlyCysLysGluSerLeuValaspGlyIleArgAlaThrAspVal 242
QY 787 ATGATCGCGCGACAGACTCTGTCATGGGTATCGCGCATGTCGCGAAGGCTCGCT 846
Db 243 MetMetAlaGlyLysValAlaValValCysGlyTyrGlyAspValGlyLysGlySerAla 262

```

```

QY 847 CAATCCCTCCGTGCCAAGCGCTCGCTTATCATCAGAACTCGACCACTGCGCT 906
Db 263 AlaSerLeuArgGlyAlaGlyAlaArgValLysValThrGluValAspProIleCysAla 282
QY 907 CTCACAGCTCCATGGAAGCTACCAAGTCCGCGCATCGAGAAAGTGGTCAAGATGTC 966
Db 283 LeuGlnAlaAlaMetAspGlyPheGluValValValLeuGluAspValAlaAspAla 302
QY 967 GATATCTTCTCATCGACGAACTCGCATATCATCTCTTCATCATGATGCGCCAG 1026
Db 303 ---IlePheIleThrThrThrGlyAsnLysAspValIleArgIleGluHisMetArgGlu 321
QY 1027 ATGAAGGATTAAGCTATTCGTTAATCGGCGCACTTCGATACGAATTCATACAGAT 1086
Db 322 MetLysAspMetAlaIleValGlyAsnIleGlyHisPheAspAsnGluIleGlnValAla 341
QY 1087 GCGCTCATGAATAACCCAGGCATCAAGCACATCCCATCAAGCCAGATACGACATGG 1146
Db 342 AlaLeuLysAsnHis-----LysTrpThrAsnIleLysAspGlnValAspMetIle 358
QY 1147 GAATTCACAGATGCCACGCTATCTCTTCTGCTGAGGCGCGCTTCTTAACTTGGC 1206
Db 359 GluMetProSerGlyAlaArgIleIleLeuLeuSerGluGlyArgLeuLeuAsnLeuGly 378
QY 1207 TCGGCTACAGTCAACCATCTTCTGTTATGTCATGTCATTCACAAACGACACTCGCT 1266
Db 379 AsnAlaThrGlyHisProSerPheValMetSerAlaSerPheThrAsnGlnValLeuAla 398
QY 1267 CAGCTCGACCTTACGAAAGAGAGAGAAATCTCGAGAAGAAGGTTTACACACTTCCGAAG 1326
Db 399 GlnIleGluLeuThrThrLysGlyAlaGluTyrGlnProGlyValTyrIleLeuProLys 418
QY 1327 CATCTCGATGAAGAAGTCGCTCCCTCCACCTCGGATCTTCGATGTCCACTTACAAAG 1386
Db 419 SerLeuAspGluLysValAlaArgLeuHisLeuLysLysIleGlyValLysLeuThrThr 438
QY 1387 CTTACACAGAGAGAGCTCACTACATCAACCTTCAGTTCAGGTCCTTACAGTCTGAT 1446
Db 439 LeuArgProAspGlnAlaGluTyrIleGlyValThrValGluGlyProPheLysSerAsp 458
QY 1447 GCTTACCGTTAT 1458
Db 459 HisTyrArgTyr 462

```

## RESULT 12

AG3505  
 C;Species: Brucella melitensis  
 C;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 15-Feb-2002  
 C;Accession: AG3505  
 R;DeiVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, .; Mazur, M.; Goltzman, E.; Seikov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letes Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
 A;Reference number: AD3252; PMID:11756688  
 A;Accession: AG3505  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-481 <KUR>  
 A;Cross-references: GB:AE008917; PIDN:AAL53210.1; PID:G17984086; GSPDB:GN00190  
 A;Experimental source: strain 16M  
 C;Genetics:  
 A;Gene: BMEI2029  
 A;Map position: 1  
 C;Superfamily: adenosylhomocysteinase  
 C;Keywords: thioether hydrolase

Alignment Scores:  
 Pred. No.: 1.21e-80 Length: 481  
 Score: 1343.00 Matches: 271  
 Percent Similarity: 70.74% Conservative: 65  
 Best Local Similarity: 57.05% Mismatches: 125



```

Query Match: 49.07% Indels: 14
DB: 2 Gaps: 4
US-09-759-990A-1 (1-1461) x AG3505 (1-481)

Qy 34 GAGTACAGAAATGCCGACATCAACCTCCATGTTCTCGCGCGTAAGCAATACCTTCCTGCT 93
Db 21 AspPheValLysAspLysLeuAlaAspTrpGlyArgGlnLeuAspLysLeuAla 40
Qy 94 GAGAGGAAATGCCAGGTCTTATGTTCTCGTGGCGTTATTCGCCCTTCTAAGCCATTG 153
Db 41 GlnThrGluMetProGlyLeuMetAlaAlaArgGlnGluPheGlyLysSerGlnProLeu 60
Qy 154 AGGGGTGCAGAACTCTGTTCCCTCCACATGACAGTCCAGACACCTCTCTATCGAG 213
Db 61 LysGlyAlaArgLysSerGlyLeuHisMetThrLysLeuAlaValLeuLeuLeu 80
Qy 214 ACACATCAGAGCTCTGTGCTGATGTCAGATGGGCTTCCTGCAACATCTTCTACACAA 273
Db 81 ThrLeuLysValLeuGlyAlaGluValArgTrpAlaSerCysAsnLysPheSerThrGln 100
Qy 274 GATACAGCGCTGCTGCTATGCTGTTGTCGCGCCCAACAGGCAACACAGAGAGCCAGCGGT 333
Db 101 AspHisAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 112
Qy 334 ATCCCGAGTCTTCGCTCGAGGCGGAAACATCTCCAGATCTCCGAGAGACACATACCGC 393
Db 113 ---ProValPheAlaValLysGlyGluThrLeuGluGluTrpTrpThrTrpAspGln 131
Qy 394 GCTCTCATGTCGCGCAGATGTCGTCGAGCGCCACAGCAGGTTGTCATGATGCTGTGATGCT 453
Db 132 IlePheGlnTrpProAspGlyLeuProSerAsnMetLysLeuAspGlyGlyAspAla 151
Qy 454 ACATCTCTATCTCCAGGCTTCGAAATTCGAAACACGCGGCTCTTCCAGAGCCAAACA 513
Db 152 ThrMetTrpLysLeuLeuLeuGlyAlaArgAlaGluAlaGlyGluAspValLeuSerAsnPro 171
Qy 514 GAAGCTGACACCTCGAATACCGTCTGCTCTCTACATCTCAAGCAGGCTCTCAACCAA 573
Db 172 GlnSerGlu-----GluGluGluValLeuPheAlaGlnLysLysArgMetAlaAla 189
Qy 574 GACAAAGAACCTGCGCACACAGTTGCTCGCGCATGAACGTTGTTCCGAGAGACACACA 633
Db 190 ThrProGlyPhePheThrLysGlnArgAlaAlaLysGlyValThrGluGluThr 209
Qy 634 ACAGTGTCCACCGCTCTACGCTCGAGAGAGAGGCGAAATCTCTCTTCCAGCCATC 693
Db 210 ThrGlyValAsnArgLeuTrpGlnLeuGlnLysGlyLeuLeuProPheProAlaIle 229
Qy 694 AACGTCAACGACGCTGTACAAAGTCCAGTTCGATTAACATCTACGCGTCCGCGCATCC 753
Db 230 AsnValAsnAspSerValThrLysSerLysPheAspAsnLysTrpGlyCysLysGluSer 249
Qy 754 CTTATCGATGTATCAACCGTCTTCGATGTCATGATCGCGCGCAAGACAGCTCTCGTC 813
Db 250 LeuValAspGlyLysArgGlyThrAspValMetAlaGlyLysValAlaValVal 269
Qy 814 ATGGGTTACGGCGATTCGCGAAGGCTCGCTCAATCTCCGTCGCGCAAGGCGCTCGC 873
Db 270 CysGlyTrpGlyAspValGlyLysGlySerAlaGlnSerLeuAlaGlyAlaValArg 289
Qy 874 GTTATCATCAGAACTCGACCAATCTGCGCTCTCCAGGCTGCCATGCAAGGCTTACCAG 933
Db 290 ValLysValThrGluValAspProLysCysAlaLeuGlnAlaAlaMetAspGlyPheGln 309
Qy 934 GTCCCGCATCGAGGAAGTCGTCAGGATGTCGATATCTTCGTTACATGACACAGGAAC 993
Db 310 ValValThrLeuAspAlaAlaSerThrAlaAspLysValValThrThrGlyAsn 329
Qy 994 TCGGATATCTCTGTTGATGATGCGCCAGATGAGAGTAAGGCTATGTCGTAAC 1053
Db 330 LysAspValLysThrLysAspHisMetArgLysMetLysAspMetCysLysValGlyAsn 349

```

---

```

Qy 1054 ATCGCCACATTCGATTAACGAATTTGATACAGATGGCTCTATGAATATACCAGGATCAAG 1113
Db 350 IleGlyHisPheAspAsnGluIleGlnValAlaAlaLeuArg-----AsnLeuLys 366
Qy 1114 CACATCCCAATCAGCCAGCAATACGACATGTGGAAATTCACAGATGGCCACGCTATCCTC 1173
Db 367 TrpThrAsnValProGlnValAspLeuLeuGlnPheProAspGlyLysArgLeuLeu 386
Qy 1174 CTTCTGCTGAGGCGCGCTTCTTAACCTTGGCTGCGCTACAGGTCAACCCATCTTTCGTT 1233
Db 387 LeuLeuSerGluGlyArgLeuLeuAsnLeuGlyAsnAlaThrGlyHisProSerPheVal 406
Qy 1234 ATGTCATGTCATTCACAAACGACACATCGCTCAGCTCGACTCTACGAAAGAGAGGA 1293
Db 407 MetSerAlaSerPheThrAsnGlnValLeuGlyLysLeuLeuPheThrArgThrAsp 426
Qy 1294 AATCTCGAGAGAGGTTTACACACTTCCGAAGCATCTCGATGATGAAGAAGTCGTCGCTC 1353
Db 427 AlaTrpLysAsnGluValTrpValLeuProLysHisLeuAspGlyLysValAlaArgLeu 446
Qy 1354 CACTCGGATCTTCGATGTCACCTTACAAAGTTCACAGAGCAGCGCTGACTACATC 1413
Db 447 HisLeuAspLysLeuGlyAlaLysLeuThrValLeuSerGluGluGlnAlaAlaLysIle 466
Qy 1414 AACGTTCCAGTTGAGGCTCTTACAGTCTGATCTTACCGTTAT 1458
Db 467 GlyValThrProGlnGlyProPheLysSerGluHisTrpArgTrp 481

```

RESULT 13  
A54040  
adenosylhomocysteinase (EC 3.3.1.1) - malaria parasite (Plasmodium falciparum)  
N;Alternate names: S-adenosyl-L-homocysteinase hydrolase  
C;Species: Plasmodium falciparum  
C;Date: 02-Aug-1994 #sequence\_revision 04-Oct-1996 #text\_change 09-Jun-2000  
C;Accession: A54040  
R;Creedon, K.A.; Rathod, P.K.; Wellems, T.E.  
J. Biol. Chem. 269, 16364-16370, 1994  
A;Title: Plasmodium falciparum S-adenosylhomocysteinase hydrolase. cDNA identification, p.  
A;Reference number: A54040; WUID:94266832; PMID:8206944  
A;Accession: A54040  
A;Molecule type: mRNA  
A;Residues: 1-479 <CRE>  
A;Cross-references: GB:U07365; NID:9537431; PIDN:AAA21391.1; PID:G460639  
C;Function:  
A;Description: catalyzes the reversible hydrolysis of S-adenosyl-L-homocysteine to aden  
C;Superfamily: adenosylhomocysteinase  
C;Keywords: NAD; thioether hydrolase  
F;258-287/Region: beta-alpha-beta NAD nucleotide-binding fold

Alignment Scores:  
Pred. No.: 6.4e-80 Length: 479  
Score: 1332.00 Matches: 257  
Percent Similarity: 69.02% Conservative: 75  
Best Local Similarity: 53.43% Mismatches: 133  
Query Match: 48.67% Indels: 16  
DB: 1 Gaps: 5

US-09-759-990A-1 (1-1461) x A54040 (1-479)

```

Qy 40 AGAATTGCGCATCAACCTCCATGTTCTCGCGCGTAAGCAACTTACCTTGTCTGAGAAG 99
Db 7 LysValLysAspLysLeuAlaProPheGlyLysMetGlnMetGlnLysSerGluAsn 26
Qy 100 GAATGCGCAGGTCTTATGTTCTTCGTGAGCGTATTCGCTTCTTAAGCCATTGAAGGT 159
Db 27 GluMetProGlyLeuMetArgLysGluGluTrpGlyLysAspGlnProLeuLysAsn 46
Qy 160 GTCAGATCTCTGTTCCCTCCACATGACAGTCCAGAGCGCTCTCTCATCGACACTC 219
Db 47 AlaLysIleThrGlyCysLeuHisMetThrValGluCysAlaLeuLeuLeuLeu 66
Qy 220 ACAGCTCTCTGCTGCTGATGTCAGATGGGCTTCCTGCAACATCTTCTCTACACAGATACA 279

```



Db 67 GlnLysLeuGlyAlaGlnIleArgTrpCysSerCysAsnIleTyrSerThrAlaAspTyr 86  
Qy 280 GCGCTGCTGTATCTGTTGTCGCCCAACAGAGCCACACAGAGAGCCAGCGGTCCCA 339  
Db 87 AlaAlaAlaVal-----SerThrLeuGlu-----AsnValThr 98  
Qy 340 GTCTTCGCTGAGAGGCGGAAACATCCACAGAACTCTGGAGAACACATACCGCGCTCTC 399  
Db 99 ValPheAlaTrpLysAsnGluThrLeuGluGluTyrTrpCysValGluSerAlaLeu 118  
Qy 400 ACATGCGCAGATGTT-----CAAGGCCACACAGCAGGTTTCGATGATGTCGTGATGCT 453  
Db 119 ThrTrpGlyAspGlyAspAsnGlyProaspMetIleValAspAspGlyGlyAspAla 138  
Qy 454 ACATCTCTCATCTCCAGGCTCTCGAA-----TTCGAAACACCGCGTCTGTT 501  
Db 139 ThrLeuLeuValHisLysGlyValGluTyrGluLysLeuTyrGluLysAsnIleLeu 158  
Qy 502 CCAGAGCCCAACAGAGCTGACACCTCCGAATACCGCTGCTTCTGCTACATCAACGAG 561  
Db 159 ProaspProGluLysAlaLysAsnGluGluGluArgCysPheLeuThrLeuLeuLysAsn 178  
Qy 562 GTCTTCAACCAAGACAGAACCACTCGCCACACAGTGTCTGCCGCATGAACGGTGTTC 621  
Db 179 SerIleLeuLysAsnProLysLysTrpThrAsnIleAlaLysLysIleIleGlyValSer 198  
Qy 622 GAAGAGACAAACAGGTGTCACCGCTCTACACGCTCGAGAGAGGCGCAACTCTCTC 681  
Db 199 GluGluThrThrThrGlyValLeuArgLeuLysLysMetAspLysGlnAsnGluLeuLeu 218  
Qy 682 TTCCAGGCATCAACGTCACAGCGCTGTACAAAGTCCAGTTCGATCAACATCTACGGC 741  
Db 219 PheThrAlaIleAsnValAsnAspAlaValThrLysGlnLysTrpAspAsnValTyrGly 238  
Qy 742 TGCAGCCATCCCTTATCGATGTATCAACCGTGTCCGATGTCATGATCGCGGCAAG 801  
Db 239 CysArgHisSerTyrThrAspGlyLeuMetArgAlaThrAspPheLeuLysSerGlyLys 258  
Qy 802 ACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 861  
Db 259 IleValValIleCysGlyTyrGlyAspValGlyLysGlyCysAlaSerMetLysGly 278  
Qy 862 CAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 921  
Db 279 LeuGlyAlaArgValTyrIleThrGluIleAspProIleCysAlaIleGlnAlaValMet 298  
Qy 922 GAAGGCTACAGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 981  
Db 299 GluGlyPheAsnValValThrLeuAspGluIleValAspLysGlyAspPheIleThr 318  
Qy 982 TGCACAGAACTCGATATCATCTCTGTTGATGATGCTGCTGCTGCTGCTGCTGCTGCT 1041  
Db 319 CysThrGlyAsnValAspValIleLysLeuGluHisLeuLeuLysMetLysAsnAla 338  
Qy 1042 ATTTCGGTAACTCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1101  
Db 339 ValValGlyAsnIleGlyHisPheAspAspGluIleGlnValAsnGluLeuPheAsnTyr 358  
Qy 1102 CAGGATCAAGCATCCCATCCATCAAGCCAGAACTGATGCTGCTGCTGCTGCTGCTGCTG 1161  
Db 359 LysGlyIleHisIleGluAsnValLysProGlnValAspArgIleThrLeuProAsnGly 378  
Qy 1162 CAGGCTATCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1221  
Db 379 AsnLysIleIleValLeuAlaArgGlyArgLeuLeuLeuLeuGlyCysAlaThrGlyHis 398  
Qy 1222 CCATCTTCTGTTATGTCAATTCACAAACACAGACACTCGCTGCTGCTGCTGCTGCTGCT 1281  
Db 399 ProAlaPheValMetSerPheSerPheCysAsnGlnThrPheAlaGlnLeuAspLeuTrp 418  
Qy 1282 GAAACAGAGA-----GGAAATCTCGAGAGAGAGTTTACACATTCGAGAGCATCTCAT 1335  
Db 419 GlnAsnLysAspThrAsnLysTyrGluAsnLysValTyrLeuLeuProLysHisLeuAsp 438

Qy 1336 GAAGAAGTCGCTCGCTCCACCTCGATCTCTCGATGCTCCACCTTACAAAGCTTACACAG 1395  
Db 439 GluLysValAlaLeuTyrHisLeuLysLysLeuAsnAlaSerLeuThrGluLeuAspAsp 458  
Qy 1396 AAGCAGGCTGACTACATCAACGTTCCAGTTGAGGTTGAGGTTGAGGTTGAGGTTGAG 1455  
Db 459 AsnGlnCysGlnPheLeuGlyValAsnLysSerGlyProPheLysSerAsnGluTyrArg 478  
Qy 1456 TAT 1458  
Db 479 Tyr 479  
RESULT 14  
A26583  
adenosylhomocysteine (EC 3.3.1.1) - rat  
N/Alternate names: S-adenosyl-L-homocysteine hydrolase  
C/Species: rattus norvegicus (Norway rat)  
C/Date: 19-Nov-1988 #sequence\_revision 04-Oct-1996 #text\_change 18-Jun-1999  
C/Accession: A26583; E26191; E26191; S69333  
R/Ogawa, H.; Gomi, T.; Mueckler, M.M.; Fujioaka, M.; Backlund Jr., P.S.; Aksamit, R.R.;  
Proc. Natl. Acad. Sci. U.S.A. 84, 719-723, 1987  
A/Title: Amino acid sequence of S-adenosyl-L-homocysteine hydrolase from rat liver as d-  
A/Reference number: A26583; MUID:87118240; PMID:3027598  
A/Accession: A26583  
A/Molecule type: mRNA  
A/Residues: 1-432 <OGA>  
A/Cross-references: GB:M15185; NID:g202803; PIDN:AAA40705.1; PID:g202804  
R/Gomi, T.; Ogawa, H.; Fujioaka, M.  
J. Biol. Chem. 261, 13422-13425, 1986  
A/Title: S-adenosylhomocysteine from rat liver. Amino acid sequences of the peptides  
A/Reference number: A26560; MUID:87008564; PMID:3759971  
A/Accession: A26191  
A/Molecule type: protein  
A/Residues: 76-94 <GOM>  
A/Accession: B26191  
A/Molecule type: protein  
A/Residues: 104-121 <G02>  
A/Note: in a manner saturably inhibited by adenosine (a product) or adenine (a competitor);  
-79 and Cys-113, suggesting a role for these residues at the active site; however, neither  
R/Merta, A.; Aksamit, R.R.; Kasir, J.; Cantoni, G.L.  
Eur. J. Biochem. 229, 575-582, 1995  
A/Title: The gene and pseudogenes of rat S-adenosyl-L-homocysteine hydrolase.  
A/Reference number: S69333; MUID:95262723; PMID:7744082  
A/Accession: S69333  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-432 <MER>  
A/Cross-references: EMBL:U14937; NID:g1223843; PIDN:AAA92043.1; PID:g1185363  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1994  
C/Genetics: 10/1, 73/3, 99/1, 149/1, 186/3, 256/1, 285/2, 324/3, 389/3  
C/Function:  
A/Description: catalyzes the reversible hydrolysis of S-adenosyl-L-homocysteine to adeno-  
C/Superfamily: adenosylhomocysteinease  
C/Keywords: homotetramer; NAD; thioether hydrolase  
F:214-243/Region: beta-alpha-beta NAD nucleotide-binding fold  
F:779,113/Active site: Cys #status predicted

## Alignment Scores:

Pred. No.:	4,24e-77	Length:	432
Score:	1289.00	Matches:	275
Percent Similarity:	66.12%	Conservative:	49
Best Local Similarity:	56.12%	Mismatches:	104
Query Match:	47.10%	Indels:	62
DB:	1	Gaps:	6

US-09-759-990A-1 (1-1461) x A26583 (1-432)

Qy 1 ATGCGTTGCCAAATCACCTACTGCTCTCCATTCGACTGACAGATTGCCGACATCAACCTC 60  
Db 1 MetAlaAspLysLeuPro-----TyrLysValAlaAspIleGlyLeu 14

QY 61 CATGTTCTCGCCGTAAGGAACCTTACCTTCTGAGGAAGAAATGACAGGCTCTTATGGTT 120  
Db 15 AlaAlaTrpGlyArgLysAlaLeuAspIleAlaGluAsnGluMetProGlyLeuMetArg 34  
QY 121 CTTCTGAGCGTATATTCGGCTTCTAAGCATTTGAAGGTTGTGAGATCTCTGTTTCCTC 180  
Db 35 MetArgGluMetTyrSerAlaSerLysProLeuLysGlyAlaArgIleAlaGlyCysLeu 54  
QY 181 CACATGACAGTCACAGACCGGCTCTCATCGAGACATCTACAGCTCTTGGTGCTGATGC 240  
Db 55 HisMetThrValGluThrAlaValLeuIleGluThrLeuValAlaLeuGlyAlaGluVal 74  
QY 241 AGATCGGCTTCTCGACATCTCTCTACACAGATACACCGCTCTCTCTATCGTGTCTC 300  
Db 75 ArgTrpSerSerCysAsnIlePheSerThrGlnAspHisAlaAlaAlaIleAla--- 93  
QY 301 GSCCCAAACAGGCACACAGAGAGCCAGCGGTTATCCAGTCTTCCCTCGGAAGGGCGAA 360  
Db 94 -----LysAlaGlyIleProValPheAlaTrpLysGlyGlu 105  
QY 361 ACATCCAGAAATCTGGGAGAACATACACCGCGCTCTCATGCGCAGATGTCACAGGC 420  
Db 106 ThrAspGluGluTyrLeuTrpCysIleGluGlnThrLeuHisPheLysAspGlyPro--- 124  
QY 421 CCACAGCAGGTGTCGATGATGTTGTGATGCTACCTCTCTCATCTCCAGGCTTCGAA 480  
Db 125 LeuAsnMetIleLeuAspAspGlyGlyAspLeuThrAsnLeuIle----- 139  
QY 481 TTCGAAACAGCGCGTGTCTCCAGAGCCAAACAGAGCTGACAACTCGAATACCGTGC 540  
Db 139 ----- 139  
QY 541 GTTCTTGCTACACTCAAGCAGGTCTTCAACACAGACAGAACCACTGGCAGACA----- 594  
Db 140 -----HisThrLysHis 143  
QY 595 -----GTTGCTCGCGCATGAACAGTGTTTCCGAGAGACAACAAGCTGTCCACGCG 648  
Db 144 ProGlnLeuLeuSerGlyIleArgGlyIleSerGluGluThrThrThrGlyValHisAsn 163  
QY 649 CTCTACCAAGCTCGAAGAGAGGGCAACTCTCTTCCAGCCATCAACGCTCAACAGCGCT 708  
Db 164 LeuTyrLysMetMetAlaAsnGlyIleLeuLysValProAlaIleAsnValAsnAspSer 183  
QY 709 GTTCAAGAGTCCAGTTTCGATACATCTACGGCTCGCGCCACTCCCTTATCGATGATC 768  
Db 184 ValThrLysSerLysPheAspAsnLeuTyrGlyCysArgGluSerLeuIleAspGlyIle 203  
QY 769 AACCGTGTCTCCGATGTCATGATCGCGGCAAGACAGCTCTCTGTCATGGTTACGGCAT 828  
Db 204 LysArgAlaThrAspValMetIleAlaGlyLysValAlaValAlaGlyTyrGlyAsp 223  
QY 829 GTCCGCAAGGCTCGCTCAATCCCTCCGTCGCGCAAGCGCTCGCTTATCATCAAGAA 888  
Db 224 ValGlyLysGlyCysAlaGlnAlaLeuArgGlyPheGlyAlaArgValIleIleThrGlu 243  
QY 889 CTCGACCAATCTCGCTCTCCAGGCTGCCATCGAGAGCTACAGGTCGCGCGCATCGAG 948  
Db 244 IleAspProIleAsnAlaLeuGlnAlaAlaMetGluGlyTyrGluValThrThrMetAsp 263  
QY 949 GAAGTCGTACAGGATGCGATATCTTCGTTATCATGACAGGAACTCGCATATCATCTCT 1008  
Db 264 GluAlaCysLysGluGlyAsnIlePheValThrThrThrGlyCysValAspIleIleLeu 283  
QY 1009 GTTGACATGATGGCCCGCATGAGGATAGAGTATTTGTCGGTAAACATCGCCCACTTCGAT 1068  
Db 284 GlyArgHisPheGluGlnMetLysAspAspAlaIleValCysAsnIleGlyHisPheAsp 303  
QY 1069 AACCAAAATGATACAGATGGCTCATGAAATACCCAGCATCAAGACATCTCCATCAAG 1128  
Db 304 ValGluIleAspValLysTrpLeuAsnGlu---AsnAlaValGluLysValAsnIleLys 322  
QY 1129 CCAGAATACGATGCGGAATTCACAGATGCGCCAGCTATCTCTCTTCTGTCAGGGC 1188

Db 323 ProGlnValAspArgTyrLeuLeuLysAsnGlyHisArgIleIleLeuAlaGluGly 342  
QY 1189 CGCTTTTAACTTCGGCTGCGCTACAGGTCAACCATCTTTTGGTTATGTCATATTCATTC 1248  
Db 343 ArgLeuValAsnLeuGlyCysAlaMetGlyHisProSerPheValMetSerAsnSerPhe 362  
QY 1249 ACAACACAGACATCGCTCAGCTCAGCTCTACGAAAGAGAGAAATCTCGAAGAAG 1308  
Db 363 ThrAsnGlnValMetAlaGlnIleGluLeuTrpThrHisProAspLysTyrProValGly 382  
QY 1309 GTTATACACATCTCCGAGCATCTCGATGAGAGCTGCTCGCTCCCTCAGCTCGATCTCTC 1368  
Db 383 ValHisPheLeuProLysLysLeuAspGluAlaValAlaGluAlaHisLeuGlyLysLeu 402  
QY 1369 GATGTCCACCTTCAAAAGCTTACACAGAGCAGCTGACTATACATCAACGTTCCAGTTGAG 1428  
Db 403 AsnValLysLeuThrLysLeuThrGluLysGlnAlaGlnTyrLeuGlyMetProIleAsn 422  
QY 1429 GGTCTTACAGTCTGATGCTTACCGTTAT 1458  
Db 423 GlyProPheLysProAspHisTyrArgTyr 432  
RESULT 15  
T40763  
adenosylhomocysteinease - fission yeast (Schizosaccharomyces pombe)  
C/Species: Schizosaccharomyces pombe  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
C/Accession: T40763  
R/Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Lauber, J.; Hilbert, H.; Duesterhoeft, A.  
submitted to the EMBL Data Library, March 1998  
A/Reference number: Z21948  
A/Accession: T40763  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-433 <LYN>  
A/Cross-references: EMBL:AL022072; PIDN:CAAI7833.1; GSPDB:GN00067, SPDB:SPBC8D2.18C  
A/Experimental source: strain 97zh-, cosmid c8D2  
C/Genetics:  
A/Gene: SPDB:SPBC8D2.18C  
A/Map position: 2  
C/Superfamily: adenosylhomocysteinease  
Alignment Scores:  
Pred. No.: 4,12e-76 Length: 433  
Score: 1274.00 Matches: 259  
Percent Similarity: 68.21% Conservative: 65  
Best Local Similarity: 54.53% Mismatches: 103  
Query Match: 46.55% Indels: 48  
DB: 2 Gaps: 5  
US-09-759-990A-1 (1-1461) x T40763 (1-433)  
QY 37 TACAGATTGCGCAGATCAACCTCCATGTTCTCGCGCTAAGGAATACCTTACCTGCTGAG 96  
Db 6 TyrLysValAlaAspIleSerLeuAlaAlaPheGlyArgLysGluLeuAlaGlu 25  
QY 97 AAGGAATGCGCAGCTTATGTTCTTCTCGAGGTTATTCGCTTCTAAGCCATGAG 156  
Db 26 AsnGluMetProGlyLeuIleAlaValArgGluLysTyrAlaLysSerGlnProLeuLys 45  
QY 157 GGTGTCAGAACTCTGTTCCCTCCACATGACAGTCCAGACAGCGCTCCATCGAGACA 216  
Db 46 GlyAlaArgIleAlaGlyCysLeuHisMetThrIleGlnThrAlaValLeuIleGluThr 65  
QY 217 CTCACAGCTCTTGGTCTGATGTCAGATGGGCTTCTCGAAACATCTTCTTACACAAGAT 276  
Db 66 LeuValAlaLeuGlyAlaGluValThrTrpSerSerCysAsnIleTyrSerThrGlnAsp 85  
QY 277 ACAGCGCTGCTGCTATGTTGTTGTCGCGCCCAACAGGCACACAGAGAACCGCGGTATC 336  
Db 86 HisAlaAlaAlaAlaIleAlaIle-----ThrGlyVal 96

```
QY 337 CCAGTCTTCGCTCGAAGGCGGAAACACTCCAGATAC---TGGGAGAACACATACCGC 393
DB 337 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 97 ProValPheAlaTyrLysGlyGluThrGluGluGluTyrLeuTrpCysIleGluGln 116
QY 394 GCTCTCATCATGCCAGATGTGTGAAGCCCAACAGCAGGTTCATCATGATGGTGTGCT 453
DB 394 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 117 LeuLysSerPheProSerGlyLysProLeuAsnMetIleLeuAspAspGlyLysPhe 136
QY 454 ACACCTCTCATCTCCAGGCTTCGAATTCGAATTCGAACAGCGGTGTCTCTCCAGACCA 513
DB 454 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 137 ThrAlaLeuValHisGluArgHis-----ProGlu----- 146
QY 514 GAAGCTGACAACTCGAATACCGCTCGCTTCTGCTACACTCAAGCAGGTCTTCAACCA 573
DB 514 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 574 GACAAGAACCACTGGCACACAGTGTCTGCGGCATGAACGGTGTTCGGAAGAGACAACA 633
DB 574 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 147 -----LeuLeuValAspIleArgGlyIleSerGluThrThr 159
QY 634 ACAGGTCTCCACCGCTCTACAGCTCGAGAGAGGAGGCAAACTCTCTTCCAGCCATC 693
DB 634 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 160 ThrGlyValHisAsnLeuTyrLysMetPheLysGluAsnLysLeuLysValProAlaIle 179
QY 694 AACCTCAAGCAGCTGTACAAAGTCCAAAGTTCGATACATCTACGGTTCGCGCCACTCC 753
DB 694 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 180 AsnValAsnAspSerValThrLysSerLysPheAspAsnLeuPheGlyCysLysGluSer 199
QY 754 CTTATCGATGGTATCAACCGTGTCTCCGATGTCATGATCGCGGCAAGACAGCTCTCGTC 813
DB 754 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 200 LeuValAspGlyIleLysArgAlaThrAspValMetIleAlaGlyLysValAlaVal 219
QY 814 ATGGGTACCGGATGTCGGAAGGCTCGCTCAATCTCCGTCGCGGCAAGCGCTCGC 873
DB 814 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 220 AlaGlyPheGlyAspValGlyLysGlyCysSerThrSerLeuArgSerGlnGlyAlaArg 239
QY 874 GTTATCATCACAGAACTCGACCAATCTCGCTCTCCAGGCTGCCATGAAGCTACCAAG 933
DB 874 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 240 ValIleValThrGluValAspProIleAsnAlaLeuGlnAlaAlaMetAspGlyPheGlu 259
QY 934 GTCGCGCGCATCGAGAGTGTCAAGGATGTCGATATCTTCGTTACATGCACAGGAAAC 993
DB 934 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 260 ValThrThrMetGluGluAlaValLysGlyGluGlnIlePheValThrThrThrGlyCys 279
QY 994 TGCATATCATCTCTGTCATGATGCGCCGATGAGGATAGGATAGGCTATTCGTTAAC 1053
DB 994 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 280 ArgAspIleIleArgGlyGluHisPheAsnGluMetLysGluAspSerIleValCysAsn 299
QY 1054 ATCGGCCCTTCGATACGAAATGATACAGATGGCTCATGAATACCCAGGCATCAAG 1113
DB 1054 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 300 IleGlyHisPheAspValGluIleAspVal---AlaTrpLeuLysAlaAsnAlaLysAsp 318
QY 1114 CACATCCCAATACAGCAGAAATACGATGTGGGAATTCACAGATGGCCACGCTATCCTC 1173
DB 1114 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 319 ValValAsnIleLysProGlnValAspArgTyrGluLeuLysAsnGlyArgHisIleIle 338
QY 1174 CTTCTTGTCTGAGGCGGCTCTTAACTTGGTGGCTACAGTACAGCTACCCATCTTCTGT 1233
DB 1174 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 339 LeuLeuAlaAspGlyArgGluValAsnLeuGlyCysAlaThrGlyHisProSerPheVal 358
QY 1234 ATGTCAATGTCAATCACAAACACACACTCGCTCGCTACAGCTCGCTTACGAAAGAGAGGA 1293
DB 1234 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 359 MetSerCysSerPheThrAsnGlnValLeuAlaGlnIleAlaLeuTrpThrAspAsnThr 378
QY 1294 AATCTCGAAGAGGTTTACACATCTCCGAGAGCATCTCGATGAAGTTCGCTCGCTC 1353
DB 1294 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 379 SerTyrProLeuGlyValHisMetLeuProLysLysLeuAspGluGluValAlaArgLeu 398
QY 1354 CACCTCGGATCTCTGATGCCACTTACAAAGCTTACACAGAGCAGGCTGCTACATC 1413
DB 1354 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 399 HisLeuGlyLysLeuGlyValLysLeuThrThrLeuThrSerValGlnSerAspTyrLeu 418
QY 1414 AACGTTCCAGTTGAGGGTCTTACAGTCTGATGCTTACCGTTAT 1458
```

Db 419 GlyIleProValAspGlyProTyrLysAlaAspHisTyrArgTyr 433

Search completed: October 15, 2003, 14:55:29  
Job time : 86 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: October 15, 2003, 14:34:29 ; Search time 30.5 seconds

(without alignments)  
4505.313 Million cell updates/sec

Title: US-09-759-990A-1

Perfect score: 2737

Sequence: 1 atggctgcaaacacactac.....ctgatgttaccgtttattaa 1461

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 255726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-O=/cn2\_1/USPTO.spool/US0975990/runat\_15102003\_115152\_2359/app\_query.fasta.1.1607  
-DB=SwissProt 41 -QPMT=fastan -SURFIX=r5p -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODES=LOCAL  
-OUTFMT=ptc -NORF=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0975990 @CN\_1\_1 37 @runat\_15102003\_115152\_2359 -NCFU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2534	92.6	486	1 SAHH_TRIVA	P51540 trichomonas
2	1645	60.1	485	1 SAHH_STRECO	Q9kzm1 streptomyce
3	1625.5	59.4	482	1 SAHH_STRAA	Q936d6 streptomyce
4	1555	56.8	471	1 SAHH_CHLIE	Q8k998 chlorobium
5	1550.5	56.6	494	1 SAHH_MYCTU	Q08364 mycobacteri
6	1544	56.4	485	1 SAHH_PETCR	Q01781 petroselinu
7	1537.5	56.2	492	1 SAHH_MYCLE	Q9ccj4 mycobacteri
8	1537	56.2	485	1 SAHH_ARATH	Q23255 arabidopsis
9	1533	56.0	485	1 SAHH_TOBAC	P50248 nicotiana t
10	1524	55.7	485	1 SAHH_WHEAT	P32112 triticum ae
11	1520	55.5	485	1 SAHH_CATRO	P35007 cactaranthu
12	1518	55.5	485	1 SAHH_MESCR	P93253 mesembryant
13	1510	55.2	480	1 SAHH_XANAC	Q9p384 xanthomonas
14	1509	55.1	485	1 SAHH_PHASS	P50249 phalaenopsi
15	1506	55.0	485	1 SAHH_LUPUL	Q9p37 lupinus lut
16	1500	54.8	480	1 SAHH_XANCP	Q9p35 xanthomonas
17	1499	54.8	485	1 SAHH_LYCES	Q9swf5 lycopersico
18	1499	54.8	485	1 SAHH_MEDSA	P50246 medicago sa

19	1468	53.6	480	1 SAHH_XYLFA	Q9pej1 xylella fas
20	1454	53.1	474	1 SAHH_CORGL	Q8nsc4 corynebacte
21	1393	50.9	474	1 SAHH_RALSO	Q9v387 ralstonia s
22	1382	50.5	466	1 SAHH_AGRIS	Q9uj99 agrobacteri
23	1374	50.2	463	1 SAHH_CAUCR	Q9abn0 caulobacter
24	1369	50.0	463	1 SAHH_RHOCA	P28183 rhodobacter
25	1359	49.7	463	1 SAHH_RHOSH	O50562 rhodobacter
26	1355	49.5	466	1 SAHH_RHIME	Q92tc1 rhizobium m
27	1349.5	49.3	466	1 SAHH_RHILO	Q98cm3 rhizobium l
28	1343	49.1	466	1 SAHH_BRUWE	Q8ve49 bruceella me
29	1337	48.8	462	1 SAHH_ROSDE	Q9znas roseobacter
30	1332	48.7	479	1 SAHH_PLAF7	P50250 plasmodium
31	1286	47.0	431	1 SAHH_RAT	P10760 rattus norv
32	1279	46.7	431	1 SAHH_MOUSE	P50247 mus musculu
33	1274	46.5	433	1 SAHH_SCHPO	Q13639 schizosacch
34	1260	46.0	432	1 SAHH_HUMAN	P23526 homo sapien
35	1252.5	45.8	432	1 SAHH_ANOGA	Q76757 anopheles g
36	1251	45.7	431	1 SAHH_DROME	Q27580 drosophila
37	1242.5	45.4	433	1 SAHH_XENLA	P51893 xenopus lae
38	1238.5	45.3	433	1 SAH2_XENLA	Q93477 xenopus lae
39	1236	45.2	437	1 SAHH_LEIDO	P36889 leishmania
40	1221.5	44.6	430	1 SAHH_DICDI	P10819 dictyosteli
41	1217.5	44.5	449	1 SAHH_YEAST	P39954 saccharomyc
42	1203.5	44.0	437	1 SAHH_CABEL	P27604 caenorhabdi
43	1142	41.7	440	1 SAHH_PNECA	Q12663 pneumocysti
44	1088.5	39.8	492	1 SAH2_DROME	P50245 drosophila
45	1061	38.8	469	1 SAHH_PSEAE	Q91685 pseudomonas

#### ALIGNMENTS

RESULT 1  
SAHH TRIVA  
ID AC P51540; STANDARD; PRT; 486 AA.  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Adenosylhomocysteinease (EC 3.3.1.1) (S-adenosyl-L-homocysteine  
DE hydrolase) (AdoCysase).  
OS Trichomonas vaginalis.  
OC Eukaryota; Parabasalidia; Trichomonadida; Trichomonadidae;  
OC Trichomonadinae; Trichomonas.  
OX NCBI\_TaxID=5722;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WAA38;  
RX MEDLINE=97047381; PubMed=8892301;  
RA Bagnara A.S., Tucker V.E., Minotto L., Howes E.R., Ko G.A.,  
RA Edwards M.R., Dawes I.W.;  
RT "Molecular characterisation of adenosylhomocysteinease from  
RT Trichomonas vaginalis.";  
RL Mol. Biochem. Parasitol. 81:11-12(1996).  
CC -!- FUNCTION: ADENOSYLHOMOCYSTEINE IS A COMPETITIVE INHIBITOR OF  
CC S-ADENOSYL-L-METHIONINE-DEPENDENT METHYL TRANSFERASE REACTIONS;  
CC THEREFORE ADENOSYLHOMOCYSTEINASE MAY PLAY A KEY ROLE IN THE  
CC CONTROL OF METHYLATIONS VIA REGULATION OF THE INTRACELLULAR  
CC CONCENTRATION OF ADENOSYLHOMOCYSTEINE.  
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine  
CC + L-homocysteine.  
CC -!- COFACTOR: NAD (BY SIMILARITY).  
CC -!- PATHWAY: Activated methyl cycle.  
CC -!- SIMILARITY: Belongs to the adenosylhomocysteinease family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; U40872; AAC47319.1; -

DR HSP; P10760; 1B3R.  
 DR InterPro; IPR000043; Ado\_hcyase.  
 DR Pfam; PF05221; AdoHcyase; 1.  
 DR Pfam; PF0670; AdoHcyase NAD; 1.  
 DR TIGRFAMs; TIGR00936; abcY; 1.  
 DR PROSITE; PS00738; ADHCYASE\_1; 1.  
 DR PROSITE; PS00739; ADHCYASE\_2; 1.  
 DR Hydrolyase; NAD; One-carbon metabolism.  
 KW NP\_BIND 267 298 NAD (POTENTIAL).  
 FT SEQUENCE 486 AA; 53433 MW; 1403C6750587427C CRC64;

Alignment Scores:  
 Pred. No.: 135e-160 Length: 486  
 Score: 2534.00 Matches: 481  
 Percent Similarity: 99.18% Conservative: 1  
 Best Local Similarity: 98.97% Mismatches: 4  
 Query Match: 92.58% Indels: 0  
 DB: 1 Gaps: 0

US-09-759-990A-1 (3-1461) x SAHH\_TRIVA (1-486)

QY 1 ATGGCTTCCAAATCACTACTGCTGCTCCATTTCGAGTACAGAAATCCGACATCAACCTC 60  
 Db 1 MetAlaCysLysSerProAlaGlyAlaProPheGluTyrArgIleAlaAspIleAsnLeu 20

QY 61 CATGTTCTCCGCCGTAAGAACTTACCCTTCTGCTGAGAGGAATGCCAGGTCTTAGGTT 120  
 Db 21 HisValLeuGlyArgLysGluLeuThrLeuAlaGluLysGluMetProGlyLeuMetVal 40

QY 121 CTTCGTGAGCGTTATTCGGCTTCTAAGCCATTGAAGGGTTCAGAAATCTCTGGTTCCTC 180  
 Db 41 LeuArgGluArgTyrSerAlaSerLysProLeuLysGlyValArgIleSerGlySerLeu 60

QY 181 CACATGACAGTCCAGACAGCGCTCTCAPCGAGACACTCAAGCTTTCTGCTGCTGATC 240  
 Db 61 HisMetThrValGlnThrAlaValLeuIleGluThrLeuThrAlaLeuGlyAlaAspVal 80

QY 241 AGATGGGCTTCTCGACANTCTCTACACAGTACAGGCTACCGCTCTCTATCTGTTGTC 300  
 Db 81 ArgTyrAlaSerCysAsnIlePheSerThrGlnAspThrAlaAlaAlaIleValVal 100

QY 301 GCGCCACAGGACACACAGAGAGAGCGCGGTATCCAGTCTTCGCTGGAAGGCGAA 360  
 Db 101 GlyProThrGlyThrProGluLysProAlaGlyIleProValPheAlaTyrLysGlyGlu 120

QY 361 ACATCCCGAATACTCGGAGAACATACCGCGCTTCTCATGCGCAGATCGTCAAGGC 420  
 Db 121 ThrLeuProGluTyrTyrGluAsnThrTyrArgAlaLeuThrTrpProAspGlyGlnGly 140

QY 421 CCACAGCAGGTGTCGATGATGCTGATGCTACCTCTCATCTCCAGGCTTCGAA 480  
 Db 141 ProGlnGlnValAlaAspAspGlyGlyAspAlaThrLeuLeuIleSerLysGlyPheGlu 160

QY 481 TTCGAAACAGCGGTGCTGCTTCCAGAGCCACACAGAGGTGACACACCTCGAATACCGCTGC 540  
 Db 161 PheGluThrAlaGlyAlaValProGluProThrGluAlaAspAsnLeuGluTyrArgCys 180

QY 541 GTTCTTGCTACACTCAGCAGGTCTTCAACACAGACAGAACACACTGGCACACAGTTGCT 600  
 Db 181 ValLeuAlaThrLeuLysGlnValPheAsnGlnAspLysAsnHisThrHisThrValAla 200

QY 601 GCGGCGATCAACGGTGTTCGAGAGACACACAGGTGTCCACCGCTCTTACCAGCTC 660  
 Db 201 AlaGlyMetAsnGlyValSerGluGluThrThrThrGlyValHisLeuGluTyrGlnLeu 220

QY 661 GAGAAGGAGGCAACCTCTCTCCAGCGCATCAACGTCACACAGCGCTGTTCACAAAGTCC 720  
 Db 221 GlnLysGluGlyLysLeuLeuPheProAlaIleAsnValAsnAspAlaValThrLysSer 240

QY 721 AGTTCGATACATCTACGGCTCGCCGCTCTCCCTTATCGATGATTCACCGCTCTTCC 780  
 Db 241 LysPheAspAsnIleTyrGlyCysArgHisSerLeuIleAspGlyIleAsnArgAlaSer 260

QY 781 GATGTCAATGATCGCGCGCAAGACAGCTCTCTCTATGCGTTACGCGCATGTTCGCAAGGCG 840  
 Db 261 AspValMetIleGlyGlyThrAlaLeuValMetGlyTyrGlyAspValGlyLysGly 280

QY 841 TCGCGTCAATCCCTCCGCTGCGCAAGCGCTCGCTTATCATCACAGAACTCGACCCCAATC 900  
 Db 281 CysAlaGlnSerLeuArgGlyGlnGlyAlaArgValIleIleThrGlnValAspProIle 300

QY 901 TCGCTCTCCAGCGTCCATGGAAGCTTACCAGTCCGCGCATCGAGGAAGTCTGTCAAG 960  
 Db 301 CysAlaLeuGlnAlaValMetGluGlyTyrGlnValArgArgIleGluValValLys 320

QY 961 GATGTCAATGATTCGTTACATGCAAGAACTCGCAATCATCTCTGTTCACATGATG 1020  
 Db 321 AspValAspIlePheValThrCysThrGlyAsnCysAspIleIleSerValAspMetMet 340

QY 1021 GCCCAGATGAAGATAAGCTATTGCTGTAACATCGGCCACTTCGATAACGAATATGAT 1080  
 Db 341 AlaGlnMetLysAspLysAlaIleValGlyAsnIleGlyHisPheAspAsnGluIleAsp 360

QY 1081 ACAGATGCTCATGAATAATACCCAGCGCATCAAGCACATCCCAATCAAGCAGAAATACGAC 1140  
 Db 361 ThrAspGlyLeuMetLysTyrProGlyIleLysHisIleProIleLysProGluTyrAsp 380

QY 1141 ATGTGGGAATCCAGATGCGCACCGCTATCTCTCTCTGCTGAGGGCGGCTTCTTAAC 1200  
 Db 381 MetTrpGluPheProAspGlyHisAlaIleLeuLeuLeuAlaGluGlyArgLeuLeuAsn 400

QY 1201 CTTCGCTCGCTCAGTCAAGTCAAGTCTTCTTATGTCAATGTCAATCACAACCAAGACA 1260  
 Db 401 LeuGlyCysAlaThrGlyHisProSerPheValMetSerMetSerPheThrAsnGlnThr 420

QY 1261 CTTCGCTCAGTTCGACCTCTACGAAAGAGAGAAATCTCGAAGAAAGTTCACACTT 1320  
 Db 421 LeuAlaGlnLeuAspLeuTyrGluLysArgGlyAsnLeuGluMetLysValTyrThrLeu 440

QY 1321 CGGAGGATCTCGATGAAGAGTCTCGCTCCACCTCGGATCTCTCGATGCTCCACTT 1380  
 Db 441 ProlLysHisLeuAspGluGluValValArgLeuHisLeuGlySerLeuAspValHisLeu 460

QY 1381 ACAAGCTTCACAGAGCAGGCTGACTACATCAACCTTCCAGTTGAGGTCCTTCAAG 1440  
 Db 461 ThrLysLeuThrGlnLysGlnAlaAspTyrIleAsnValProValGluGlyProTyrLys 480

QY 1441 TCTGATGCTTACCGTTAT 1458  
 Db 481 SerAspAlaTyrArgTyr 486

RESULT 2  
 SAHH\_STRCO STANDARD; PRT; 485 AA.  
 ID SAHH\_STRCO Q9KZM1;  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 42, Last annotation update)  
 DE Adenosylhomocysteinease (SC 3.3.1.1) (S-adenosyl-L-homocysteine  
 DE hydrolase) (AdoHcyase).  
 GN AHY OR SAHH OR SC03023 OR SCE34.04C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RC MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kisser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kisser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,

RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RT Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 CC -|- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine  
 CC + L-homocysteine.  
 CC -|- COFACTOR: NAD (By similarity).  
 CC -|- PATHWAY: Activated methyl cycle.  
 CC -|- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -|- SIMILARITY: belongs to the adenosylhomocysteinase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL939114; CAB8907.1; -;  
 DR HSSP; P10760; 1D4F.  
 DR HAMAP; MF 00563; -; 1.  
 DR InterPro; IPR000043; Ado\_hcyase.  
 DR Pfam; PF05221; AdoHcyase; 1.  
 DR Pfam; PF00670; AdoHcyase\_NAD; 1.  
 DR TIGRFAMs; TIGR00936; ahcy; 1.  
 DR PROSITE; PS00738; ADOHCYASE 1; 1.  
 DR PROSITE; PS00739; ADOHCYASE 2; 1.  
 KW Hydrolase; NAD; One-carbon metabolism; Complete proteome.  
 FT NP\_BIND 266 297 NAD (POTENTIAL).  
 SQ SEQUENCE 485 AA; 52940 MW; 45ASC0C9175C81F7 CRC64;

## Alignment Scores:

Pred. No.: 1,288-1,01 Length: 485  
 Score: 1645.00  
 Percent Similarity: 78.83%  
 Best Local Similarity: 65.62%  
 Query Match: 60.10%  
 DB: 1  
 Gaps: 1

US-09-759-990A-1 (1-1461) x SAHH\_STRCO (1-485)

QY 34 GAGTACAGAAATCCGACATCAACCTCCATGTTCTCGCGCTAAGGACATCACTTCCTGCT 93  
 Db 9 ASPPhelysValAlaAspLeuSerLeuAlaAlaPheGlyArgLysGluIleThrLeuAla 28  
 QY 94 GAGAGGAAATCCAGGCTTATGTTCTTCGTGAGCGTTATCCGCTTCTAAGCCATTG 153  
 Db 29 GluHisGluMetProGlyLeuMetAlaIleArgLysGluTyrAlaGluAlaGlnProLeu 48  
 QY 154 AAGGTTGTCAGATCTCTGTTCCCTCCACATGACAGTCCAGACAGCGCTCTCATCGAG 213  
 Db 49 AlaGlyAlaArgValThrGlySerLeuHisMetThrValGlnThrAlaValLeuIleGlu 68  
 QY 214 ACACCTCACAGCTCTTGCTGCTGATGTCAGATGGGCTTCTCGAACATCTTCTCTACAAA 273  
 Db 69 ThrLeuValAlaLeuGlyAlaGluValArgTyrAlaSerCysAsnIlePheSerThrGln 88  
 QY 274 GATACAGCGCTGCTGCTATCGTTGTCGCGCCAAAGGACACACAGGACAGCCGCGCT 333  
 Db 89 AsphiAlaAlaAlaAlaAlaIleAlaValGlyProAsnGlyThrProAspAsnProGlnGly 108  
 QY 334 ATCCCGATCTTCGCTGAGGCGGAACACTCCAGATACTCGGAGACACATACCGC 393  
 Db 109 ValProValPheAlaTrpGlyGluThrLeuGluGluTyrTrpTrpCysThrGluGln 128  
 QY 394 GCTCTCACATGGCCAGATGGTCAA-----GGCCACACAGAGTTGTGATGATGGTGT 447  
 Db 129 AlaLeuThrTrpProAsnThrProThrGlyGlyProAsnMetIleLeuAspAspGlyGly 148  
 QY 448 GATGCTACATCTCTCAAGGGCTTCGAATTCGAACACAGCGGCTGTTCAGAG 507

Db 149 AspaIaThrLeuLeuValHisLysGlyValGluTyrGluLysAspGlyLysValProSer 168  
 QY 508 CCAACAGAACTGACAACTCGAATACCGTTCGCTCTTCTGCTACACTCAAGCAGGTCTTC 567  
 Db 169 ValAspThrAlaGluSerAspGluHisArgValIleLeuLeuLeuLeuThrArgThrVal 188  
 QY 568 AACCAAGAACAGAACCACTGGCACACAGTGTCTGCGGCGATGAACGGTGTTCGGAAGAG 627  
 Db 189 GlyGluSerProGlnLysTrpThrGlnLeuAlaSerGluIleArgGlyValThrGluGlu 208  
 QY 628 ACACAAACAGGTCTCCACCGCTCTTACAGCTCGAAGAGGAGGCAAACTCTCTCCCA 687  
 Db 209 ThrThrThrGlyValHisArgLeuTyrGluMetHisArgAspGlyThrLeuLeuPhePro 228  
 QY 688 GCCATCAACCTCAACGACCTCTTACAAAGTCCAAAGTTCGATAACATCTACGGTCCGCG 747  
 Db 229 AlaIleAsnValAsnAspAlaValThrLysSerLysPheAspAsnLysTyrGlyCysArg 248  
 QY 748 CACTCCCTTATCGATGTTATCAACCGTCTCCGATGTCATGATCGCGGCAAGCAGCT 807  
 Db 249 HisSerLeuIleAspGlyIleAsnArgAlaThrAspValLeuIleGlyLysThrAla 268  
 QY 808 CTGCTCATGGTTACGGCGATCTCGCAAGGGTGGGCTCAATCCCTCCGTGGCCAGGC 867  
 Db 269 ValValCysGlyTyrGlyAspValGlyLysGlyCysAlaGluSerLeuArgGlyGlnGly 288  
 QY 868 GCTCGGTTATCATACAGAACTCGACCAATCTGGGCTCTCCAGGCTCCATGGAAGGC 927  
 Db 289 AlaArgValIleIleThrGluIleAspProIleCysAlaLeuGlnAlaAlaMetAspGly 308  
 QY 928 TACCAGGTCCCGCGATCCAGGAAGTCTCTCAAGATGTCGATATCTTCTGTTACATGCACA 987  
 Db 309 PheGlnValThrThrLeuAspGluValValAspLysAlaAspIlePheValThrThr 328  
 QY 988 GGAACCTGCGATATCATCTCTTGCATGATGCGCCAGATGAAGATAAGCTATTGTC 1047  
 Db 329 GlyAsnLysAspIleIleMetAlaLysAspMetAlaLysMetLysHisGlnAlaIleVal 348  
 QY 1048 GGTACATCGCCACTTCGATACGAATTCGATACAGATGGCTCATGAATACCCAGGC 1107  
 Db 349 GlyAsnIleGlyHisPheAspAsnGluIleAspMetAlaGlyLeuAlaGlnThrProGly 368  
 QY 1108 ATCAAGCACATCCCAATCAAGCCAGAAATACGACATGTGGGAATTCCTCCAGTGGCCAGCT 1167  
 Db 369 IleValLysAspGluValLysProGlnValHisThrTrpThrTyrProAspGlyLysVal 388  
 QY 1168 ATCTCTCTTCTGTCAGGCGCGCTTCTTAACTTGGCTGCGCTACAGTCAACCTCTCT 1227  
 Db 389 LeuIleValLeuSerGluGlyArgLeuLeuAsnLeuGlyAsnAlaThrGlyHisProSer 408  
 QY 1228 TTCTGTTATGTCATGTCTATTCAAAACACAGACACTCGCTCAGCTCGACTTACGAAAG 1287  
 Db 409 PheValMetSerAsnSerPheAlaAspGlnThrLeuAlaGlnIleGluLeuPheThrLys 428  
 QY 1288 AGAGGAATCTCGAAGAGAGTTTACACACTCCGAGCATCTCGATGAAGAAGTCTGCT 1347  
 Db 429 ProAspGluTyrProThrAspValTyrValLeuProLysHisLeuAspGluLysValAla 448  
 QY 1348 CCGCTCCACCTCGGATCTCTCGATGTCACCTTACAAAGCTTACACAGACAGCTGAC 1407  
 Db 449 ArgLeuHisLeuAspSerLeuGlyValLysLeuThrThrLeuArgProGluGlnAlaAsp 468  
 QY 1408 TACATCAACGTTCCAGTTCAGGCTCTTACAAAGTCTGATGCTTACCGTTAT 1458  
 Db 469 TyrIleGlyValLysValGluGlyProTyrLysAlaAspHisTyrArgTyr 485

## RESULT 3

SAHH\_STRAA  
 ID SAHH\_STRAA STANDARD; PRT; 482 AA.  
 AC Q936D6;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)



RESULT 4  
SAHH CHLITE STANDARD; PRT; 471 AA.  
AC O8KEG8;  
AT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Adenosylhomocysteinease (EC 3.3.1.1) (S-adenosyl-L-homocysteine  
DE hydrolase) (AdoHcyase).  
GN AHY OR SAHH OR CT0721.  
OS Chlorobium tepidum.  
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;  
OC Chlorobium.  
OX NCBI\_TaxID=1097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TLS / ATCC 49652 / DSM 12025;  
RX MEDLINE=22103685; PubMed=12093101;  
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.P., Wu M.,  
RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,  
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,  
RA Holt I., Unayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,  
RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,  
RA Vamathevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,  
RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;  
RA "The complete genome sequence of Chlorobium tepidum TLS, a  
RT photosynthetic, anaerobic, green-sulfur bacterium.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).  
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine  
CC + L-homocysteine.  
CC -1- COFACTOR: NAD (By similarity).  
CC -1- PATHWAY: Activated methyl cycle.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -1- SIMILARITY: Belongs to the adenosylhomocysteinease family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AE012843; HAM71958.1; -  
DR TIGR: CT0721; -; 1.  
DR HAVAP: MF\_00563; -; 1.  
DR InterPro: IPR000043; Ado\_hcyase.  
DR InterPro: IPR000205; NAD\_binding.  
DR Pfam: PF05221; AdoHcyase; 1.  
DR Pfam: PF00670; AdoHcyase; NAD; 1.  
DR TIGRFAMS: TIGR00936; ahcy\_1.  
DR PROSITE: PS00738; ADHCYASE\_1; 1.  
DR PROSITE: PS00739; ADHCYASE\_2; 1.  
FW Hydrolase; NAD; One-carbon metabolism; Complete proteome.  
KT NP BIND 254 285 NAD (POTENTIAL).  
SQ SEQUENCE 471 AA; 51949 MW; 31B082405D241047 CRC64;  
Alignment Scores:  
Pred. No.: 1.18e-95 Length: 471  
Score: 1555.00 Matches: 303  
Percent Similarity: 76.36% Conservative: 62  
Best Local Similarity: 63.39% Mismatches: 101  
Query Match: 56.81% Indels: 12  
DB: 1 Gaps: 3  
US-09-759-990A-1 (1-1461) x SAHH\_CHLITE (1-471)  
QY 25 GTCCTCCAGTACAGATTCGCGACATCAACCTCCATCTTCCTCGCGCGTAAGCACTT 84  
Dd 6 AValLeuAspTyrLysValAlaAspIleSerLeuAlaGluTrpGlyArgLysGluLe 25  
QY 85 ACCCTTGCTGAGAGGAATGCCAGGTCTTATGTTCTTCGTGAGCGTTATTCGCTTCT 144

Dd 26 GluIleAlaGluLysGluMetProGlyLeuMetAlaThrArgLysLysTyrGluGlyLys 45  
QY 145 AAGCCATTGAAGGGTGTCAAGATCTCTGTTCCCTCCACATGACAGTCCAGACACCGTC 204  
Dd 46 LysProLeuAlaGlyAlaArgIleAlaGlySerLeuHisMetThrIleGlnThraVal 65  
QY 205 CTCATCGACACATCAAGCTCTTGTGCTCATGATCGATGCGCTTCCTGCAACATCTTC 264  
Dd 66 LeuIleGluThrLeuValGluLeuGlyAlaAspValArgTrpAlaSerCysAsnIlePhe 85  
QY 265 TCTACACAGATACAGCGCTGCTGCTTCTGCTTCTGCGCCCAACAGCACACAGAGAG 324  
Dd 86 SerThrGlnAspHisAlaAlaIleAla----- 97  
QY 325 CCAGCGGTATCCCGATCTTCGCTCGAAGGGCCAAACACTCCAGAAATCTGGGAGAAC 384  
Dd 98 ---AlaGlyValProValPheAlaTrpLysGlyGluThrLeuAspGluTyrTrpTrpCys 116  
QY 385 ACATACCGCGCTCTCATGCGCCAGATGTCAGGCCACAGCAGGTTGTCGATGATGTT 444  
Dd 117 ThrArgGlnIleLeuGluPheGluGlyLeuGlyProAsnLeuIleValAspAspGly 136  
QY 445 GGTGATGCTACACTCTCTCTCCAAAGGGCTTCAATTCGAAACACGCGGTGCTGCCA 504  
Dd 137 GlyAspAlaThrLeuMetIleHisPheGlyTyrLysIleGluAsnAspProSerMetLeu 156  
QY 505 GAGCCACAGAGAGTCAACACTCGAATACCGCTTCTTCTTCTACACTCAAGCAGGTC 564  
Dd 157 AspLysThrProGly--AsnAlaGluLysAlaLeuLeuGlnIleLeuLysAlaVal 175  
QY 565 TTCACACACAGAGAACACTGGCACACAGTCTCTCGCGCATGAAACGCTGTTCCGAA 624  
Dd 176 PheAlaGluAspAsnGlnArgTrpHisLysValAlaAlaGlyMetLysGlyValSerGlu 195  
QY 625 GAGACACAAACAGGTGTCCACCGCTCTACAGAGTCCAGAGGAGGAGGAACTCTCTTC 684  
Dd 196 GluThrThrGlyValHisArgLeuTyrGlnMetLysGlyGluLeuPhe 215  
QY 685 CAGGCATCAAGTCACAGACGCTGTTCAAAAGTCCAAAGTTCGATAACATCTACGGCTGC 744  
Dd 216 ProAlaIleAsnValAsnAspSerValThrLysSerLysPheAspAsnLeuTyrGlyCys 235  
QY 745 CGCCACTCCCTTATCATGATCAACCGTCTCCATGTCATGATCGCGGCAAGACA 804  
Dd 236 ArgGluSerLeuAlaAspGlyIleLysArgAlaThrAspValMetIleAlaGlyLysVal 255  
QY 805 GCTCTCGTCATGGTTACGCGCATGTCGCAAGGCTGCGCTCAATCCCTCGTGGCAA 864  
Dd 256 ValValValLeuGlyTyrGlyAspValGlyLysGlyCysAlaHisSerMetArgSerTyr 275  
QY 865 GCGCTCGCTTATCATCAAGAACTCGAACCAATCTGCGCTCTCAGGCTGCCATGAA 924  
Dd 276 GlyAlaArgValIleValThrGluIleAspProIleCysAlaLeuGlnAlaAlaMetGlu 295  
QY 925 GGCATACAGGTCCCGCATCGAGGAAGTCTCAAGATGTCGATATCTTCGTACATGC 984  
Dd 296 GlyPheGluValThrThrMetGluAlaValLysGluIlePheValThrAla 315  
QY 985 ACAGGAAACTCGCATATCATCTCTGTGACATGAGGCCCGCATGAAGTAAGGCTATT 1044  
Dd 316 ThrGlyAsnLysAspValIleThrLeuAspHisIleLysGlnMetArgAspGluAlaIle 335  
QY 1045 GTCGGTACATTCGGCCACTTCGATTAACGAATTCGATGAGATGGCTCATGAATACCCA 1104  
Dd 336 ValCysAsnIleGlyHisPheAspAsnGluIleGlnValAspAlaLeuAsnAsnPheLys 355  
QY 1105 GGCATCAAGCACATCCCAATCAAGCCAGAAATCAAGATGTCGGAATTCAGATGCGCAC 1164  
Dd 356 GlyAlaThrArgIleAsnIleLysProGlnValAspLysTyrValPheGluAsnGlyAsn 375  
QY 1165 GCTATCTCTCTTCTGCTGAGGCGCGCTTCTTAACCTTGGCTGGCTACAGGTCAACCA 1224



Db 376 CysleTyrlleuLeuAlaGluGlyArgleuValasleuGlyCysAlaThrChlyHisPro 395  
 QY 1225 TCTTTCTGTAATGATGATCATTCACAAACGAGCATCTCGCTCAGTCGACCTTACGAA 1284  
 Db 396 SerPheValMetSerAsnSerPheThrAsnGlnThrLeuAlaGlnleuLeuTrpGln 415  
 QY 1285 AAGAGAGAAATCTCGAGAAGAAGGTTTACACACTTCCGAAGCATCTCGATGAAGAAGTC 1344  
 Db 416 -----AsnAspTyrlsValGlyValTyrlsValGlyValTyrlsValGlyVal 433  
 QY 1345 GTCGCTCCACTCGCATCTCTCGATGTCACACCTTACAAAGCTTACACAGAGAGCGCT 1404  
 Db 434 AlaArgleuHisleuGlyGlnleuGlyAlaLysleuThrThrLeuThrlysluGlnAla 453  
 QY 1405 GACTACATCAAGTTCAGTTCAGGTCCTTACAGTCTGATCTACCGTTAT 1458  
 Db 454 AspTyrlleuGlyValProValGluGlyProTyrlsProGluHisTyrlsArgTyr 471  
 RESULT 5  
 SAHH MYCTU  
 ID SAHH MYCTU STANDARD; PRT; 494 AA.  
 AC 008364; P81859;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Adenosylhomocysteine (EC 3.3.1.1) (S-adenosyl-L-homocysteine  
 DE hydrolase) (AdoHcyase).  
 GN AHCY OR SAHH OR RV3248C OR WT3346 OR MTCY20B11.23C.  
 OS Mycobacterium tuberculosis, and  
 OS Mycobacterium bovis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773, 1765;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares J.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.P., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RA "whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=M.bovis; STRAIN=BCG / Pasteur;  
 RA Pawar S.N., Nayak R.;  
 RT "Complete genomic DNA sequence of the S-adenosyl-L-homocysteine  
 RT hydrolase gene from Mycobacterium bovis BCG."  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 1-19 AND 225-245.  
 RC SPECIES=M.bovis; STRAIN=BCG / Pasteur;  
 RA Pawar S.N., Balaji K.N., Jaffe H., Nayak R.;  
 RL Submitted (MAY-1999) to the Swiss-Prot data bank.  
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine  
 CC + L-homocysteine.

CC -1- COFACTOR: NAD.  
 CC -1- PATHWAY: Activated methyl cycle.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: Belongs to the adenosylhomocysteine family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z95121; CAB08349.1; -;  
 DR EMBL; AE007145; AAK47688.1; -;  
 DR EMBL; AF262755; AAF72670.1; -;  
 DR PIR; B70593; B70593.  
 DR HSP; P10760; 1D4F.  
 DR TIGR; MT3346; -;  
 DR Tuberculist; RV3248c; -;  
 DR HAMAP; MF\_00563; -; 1.  
 DR InterPro; IPR000043; Ado\_hcyase.  
 DR Pfam; PF05221; AdoHcyase; 1.  
 DR Pfam; PF0670; AdoHcyase\_NAD; 1.  
 DR TIGRFAMs; TIGR00936; ahcY; 1.  
 DR PROSITE; PS00738; ADHCHYASE\_1; 1.  
 DR PROSITE; PS00739; ADHCHYASE\_2; 1.  
 KW Hydrolase; NAD; One-carbon metabolism; Complete proteome.  
 FT INIT MET 0  
 FT NP BIND 275 306 NAD (POTENTIAL).  
 SQ SEQUENCE 494 AA; 54192 MW; 2A27D6002B2AFB16 CRC64;  
 Alignment Scores:  
 Pred. No.: 2,38e-95 Length: 494  
 Score: 1550.50 Matches: 297  
 Percent Similarity: 76.52% Conservative: 68  
 Best Local Similarity: 62.26% Mismatches: 109  
 Query Match: 56.65% Indels: 3  
 DB: 1 Gaps: 3  
 US-09-759-990A-1 (1-1461) x SAHH\_MYCTU (1-494)  
 QY 34 GAGTACAGATTCGCGACATCACTCCATGTTCTCGCGCTAAGCACTTACCTTGCT 93  
 Db 19 AspPheLysleuAlaAspSerLeuAlaAspPheGlyArgLysGluLeuArgileAla 38  
 QY 94 GAGAAGAAATGCGAGTCTTATGTTCTTCGTGAGCGTTATTCGCTTCTAAGCCATTG 153  
 Db 39 GluHisGluMetProGlyLeuMetSerLeuArgGluTyrlaGluValGlnProleu 58  
 QY 154 AAGGTGTCAATCTCTGTTCCCTCCATGACAGTCCAGACGCCGCTCTCATCGAG 213  
 Db 59 LysGlyAlaAlaArgileSerGlySerLeuHisMetThrValGlnThrAlaValleuIleGlu 78  
 QY 214 ACATCTCAGCTTCTGTCGTCTGATGTCAGATGGCTTCCTGCAACATCTTCTTACCAA 273  
 Db 79 ThrLeuThrAlaLeuGlyAlaGluValArgTrpAlaSerCysAsnIlePheSerThrGln 98  
 QY 274 GATACAGCCGCTGCTGCTATCGTTCTCGGCCCAACAGACACACAGAACGCCGCGT 333  
 Db 99 AsphilaAlaAlaAlaValValGlyProHisGlyThrProAspGluProLysGly 118  
 QY 334 ATCCCATCTTCCCTGGAAGGCGAACAACACTCCCGAATATCTGGGAGAACACATACCGC 393  
 Db 119 ValProValPheAlaTrpLysGlyGluThrLeuGluGluTyrlTrpAlaAlaGluGln 138  
 QY 394 GCTCTCATGCGCCATGTCAGCCCA---CAGCAGCTTGTGATGATGCTGTCAT 450  
 Db 139 MetLeuThrTrpProAspProAspLysProAlaAsnMetLeuAspGlyGlyAsp 158  
 QY 451 GCTACACTCTCTCATCTCCAAAGGGCTTCGAATTCGAACACAGCCGCTGTTCCAGACCA 510  
 Db 159 AlaThrMetLeuValleuArgGlyMetGlnTyrlGluLysAlaGlyValValProAla 178

```

QY 511 ACAGAGCTGACACCTCGCATACCGCTCGCTTCTTGCTACATCAAGCAGGTCTTCAAC 570
Db 179 GluGluAspAspProAlaGluThrPheValPheLeuAsnLeuLeuArgThrArgPheGlu 198
QY 571 CAAGACAAAGAACCACTCGCACACAGTGTGCGCGCATGAACGGTGTTCACGAAGACACA 630
Db 199 ThrAspLysAspLysTyrThrLysIleAlaGluSerValLysGlyValThrGluGluThr 218
QY 631 ACAACAGGTGTCACCGCTCTACAGCTCGAGAGAGGAGCAACTCTCTTCCACGCC 690
Db 219 ThrThrGlyValLeuArgLeuTyrGlnPheAlaAlaGlyAspLeuAlaPheProAla 238
QY 691 ATCAACGTCACAGCGCTGTACAAAGTCCAGTTCGATACATCACTACCGTCCGCCAC 750
Db 239 IleAsnValAsnAspSerValThrLysSerLysPheAspAsnLysTyrGlyThrArgHis 258
QY 751 TCCTTATCGATGTATCAACCGTCTCCGATGTCATGTCGCGCGCAAGACAGCTCTC 810
Db 259 SerLeuIleAspGlyIleAsnArgGlyThrAspAlaLeuIleGlyLysLysValLeu 278
QY 811 GTCATGGTTACGGCGATGTCGCAAGGCTCGCTCAATCCCTCCGTCGCCAAGCGCT 870
Db 279 IleCysGlyTyrGlyAspValGlyLysGlyCysAlaGluAlaMetLysGlyGlnGlyAla 298
QY 871 CGCGTATCATCAGACACTCGACCAATCTCGCTCTCCAGCTGCCATGCAAGGCTAC 930
Db 299 ArgValSerValThrGluIleAspProIleAsnAlaLeuGlnAlaMetMetGluGlyPhe 318
QY 931 CAGGTCCCGCATCGAGGAAGTCGTCAGGATGTCATATCTTCTGTCATGCACAGGA 990
Db 319 AspValValThrValGluGluAlaIleGlyAspAlaAspIleValValThrAlaThrGly 338
QY 991 ACTCGCATATCATCTCTGTGCATGATGATGCGCCAGATGAGGATAGGCTATTCTCGT 1050
Db 339 AsnLysAspIleIleMetLeuGluHisIleLysAlaMetLysAspHisAlaIleLeuGly 358
QY 1051 ACATCGCCACTTCGATACGAAATGATACAGATGCGCTCATGAAATACCCAGCATC 1110
Db 359 AsnIleGlyHisPheAspAsnGluIleAspMetAlaGlyLeuGluArg---SerGlyAla 377
QY 1111 AAGCATATCCCAATCAAGACAGAAATACGATGTCGGAATTCACAGAT---GCCACGCT 1167
Db 378 ThrArgValAsnValLysProGlnValAspLeuTyrThrPheGlyAspThrGlyArgSer 397
QY 1168 ATCTCTCTTCTGTCAGGCGCGCTTCTTAACCTTGCTGCGCTACAGCTCACCATCT 1227
Db 398 IleIleValLeuSerGluGlyArgLeuLeuAsnLeuGlyAsnAlaThrGlyHisProSer 417
QY 1228 TTCTGTTATGTCATGTCATTCACAAACAGACACTCGCTCAGCTCGACTCTACGAAG 1287
Db 418 PheValMetSerAsnSerPheAlaAsnGlnThrIleAlaGlnIleGluLeuTyrThrLys 437
QY 1288 AGAGAAATCTCGAAGAAGGTTTACACATCTCCGAAGCATCTCGATCAAGAAGTCGCT 1347
Db 438 AsnAspGluTyrAspAsnGluValTyrArgLeuProLysHisLeuAspGluLysValAla 457
QY 1348 CGCTCCACCTCGGATCTCTCGATGTCACCTTACAAAGCTTACAGAAAGCAGGCTGAC 1407
Db 458 ArgIleHisValGluAlaLeuGlyGlyHisLeuThrLysLeuThrLysGluGlnAlaGlu 477
QY 1408 TACATCAAGTTCAGTGGCTGCTTACAGTCTGATGCTTACCGTTAT 1458
Db 478 TyrLeuGlyValAspValGluGlyProTyrLysProAspHisTyrArgTyr 494

```

RESULT 6

SAHH\_PETCR

ID -SAHH\_PETCR STANDARD; PRT; 485 AA.

AC Q01781;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Adenosylhomocysteinase (EC 3.3.1.1) (S-adenosyl-L-homocysteine

```

DE hydrolase) (AdoHcyase).
GN SAHH OR SHH.
OS Petroselinum crispum (Parsley) (Petroselinum hortense).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Apiales; Apiaceae; Petroselinum.
OX NCBI_TaxID=4043;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=92262510; PubMed=1374911;
RX Kowaleck P., Plesch G., Hahlbrock K., Somssich I.E.;
RA "Induction by fungal elicitor of S-adenosyl-L-methionine synthetase
RT and S-adenosyl-L-homocysteine hydrolase mRNAs in cultured cells and
RL leaves of Petroselinum crispum."
RL Proc. Natl. Acad. Sci. U.S.A. 89:4713-4717 (1992).
RN [2]
RP SEQUENCE OF 259-485 FROM N.A.
RA Somssich I.E., Bollmann J., Hahlbrock K., Kombrink E., Schulz W.;
RT "Differential early activation of defense-related genes in
RT elicitor-treated parsley cells."
RL Plant Mol. Biol. 12:227-234 (1989).
CC -|- FUNCTION: ADENOSYLHOMOCYSTEINE IS A COMPETITIVE INHIBITOR OF
CC S-ADENOSYL-L-METHIONINE-DEPENDENT METHYL TRANSFERASE REACTIONS;
CC THEREFORE ADENOSYLHOMOCYSTEINE MAY PLAY A KEY ROLE IN THE
CC CONTROL OF METHYLATIONS VIA REGULATION OF THE INTRACELLULAR
CC CONCENTRATION OF ADENOSYLHOMOCYSTEINE.
CC -|- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine
CC + L-homocysteine.
CC -|- COFACTOR: NAD.
CC -|- PATHWAY: Activated methyl cycle.
CC -|- SUBUNIT: Homotetramer.
CC -|- TISSUE SPECIFICITY: MAINLY IN FLORAL BUDS AND STEMS.
CC -|- SIMILARITY: Belongs to the adenosylhomocysteinase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; M81885; AAA33856.1; -.
CC EMBL; M62756; AAA33855.1; -.
CC PIR; T15035; T15035.
CC HSP; P10760; I33R.
CC InterPro; IPR000043; Ado_hcyase.
CC Pfam; PF05221; AdoHcyase_1.
CC Pfam; PF00670; AdoHcyase_NAD; 1.
CC TIGRfam; TIGR00936; ahcy; 1.
CC PROSITE; PS00738; ADOHCYASE_1; 1.
CC PROSITE; PS00739; ADOHCYASE_2; 1.
CC Hydrolase; NAD; One-carbon metabolism.
CC NP_BIND; 263 294 NAD (POTENTIAL).
CC FT CONFLICT 346 346 C -> D (IN REF. 1; AAA33855).
CC FT CONFLICT 439 439 L -> C (IN REF. 1; AAA33855).
CC SQ SEQUENCE 485 AA; 53181 MW; 05E926516C2B08E6 CRC64;

```

Alignment Scores:

Pred. No.:	6 39e-95	Length:	485
Score:	1544.00	Matches:	308
Percent Similarity:	74.59%	Conservative:	56
Best Local Similarity:	63.11%	Mismatches:	106
Query Match:	56.41%	Indels:	18
DB:	1	Gaps:	6

US-09-759-990A-1 (1-1461) x SAHH\_PETCR (1-485)

QY 19 ACTGTCCTCCATTCGATACAGAAATCCGACATCAACCTTCATGTCGCGCGTAAG 78

Db 8 ThrAlaAlaGlyArgGluTyrLysValLysPheMetSerLeuAlaAspPheGlyArgLeu 27

Db	378	TrpValIpeProAspThrGlyA-ggIyIleIleIleLeuAlaGluGlyArgIeuMetAsn	397
QY	1201	CTTGGCGTGGCTACAGGFTCCACCATCTTTCGGTATGTCAATGTCAATTCACAAACAGACACA	1260
Db	398	LeuGlyCYeAlaThrGlyHisProSerPheValMetSerCysSerPheThrAsnGlnVal	417
QY	1261	CTGGCTCACTCGACTTCAGTCTACGAAAGAGA-----GGAATCTCGAAGAGAGGTTTAC	1314
Db	418	ILeAlaGlnLeuGluLeuTrpAsnGluLysSerSerGlyIysTrpGluLysLysValTyr	437
QY	1315	ACACTCCGAACATCTCGATGAAGAAGTCGGCTCGCTCCACCTCCGATCTCTCGATGTC	1374
Db	438	ValIeuProLysHisLeuAspGluLysValAlaAlaLeuHisLeuGlyLysLeuGlyAla	457
QY	1375	CACCTTACAAAGCTTTACACAGACAGGCTGACTCATCAACAGTTCAGGTTAGGGTCTCT	1434
Db	458	LysLeuThrLysLeuSerLysAspGlnAlaAspTyrIleSerValProValGluGlyPro	477
QY	1435	TACAGACTGCTGATGCTTACCGTTAT	1458
Db	478	TyrLysProAlaHisTyrArgTyr	485
RESULT 7			
SAHH MYCLE			
ID	SAHH MYCLE	STANDARD;	PRT; 492 AA.
AC	Q9CCJ4;		
DT	28-FEB-2003	(Rel. 41, Created)	
DT	28-FEB-2003	(Rel. 41, Last sequence update)	
DT	28-FEB-2003	(Rel. 41, Last annotation update)	
DE	Adenosylhomocysteine (BC 3.3.1.1) (S-adenosyl-L-homocysteine hydrolase) (AdoHcyase).		
DE	AHCY OR SAHH OR ML0771.		
OS	Mycobacterium leprae.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.		
ON	NCBI_TaxID=1769;		
OX	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=TN;		
RX	MEDLINE=21128732; PubMed=11234002;		
RA	Cole S.T., Bigmeier K., Parkhill J., James K.D., Thomson M.R.,		
RA	Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,		
RA	Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,		
RA	Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,		
RA	Hollroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,		
RA	Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,		
RA	Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,		
RA	Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,		
RA	Barrell B.G.;		
RT	"Massive gene decay in the leprosy bacillus.";		
RL	Nature 409:1007-1011(2001).		
CC	-!- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine + L-homocysteine.		
CC	-!- COFACTOR: NAD (By similarity).		
CC	-!- PATHWAY: Activated methyl cycle.		
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).		
CC	-!- SIMILARITY: Belongs to the adenosylhomocysteine family.		
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	EMBL; AL583919; CAC30280.1; -		
DR	PIR; D87005; D87005.		
DR	HSP; P10760; ID4P.		
DR	Leprama; ML0771; -.		
DR	HMAP; MF_0563; -; 1.		
DR	InterPro; IPR000043; Ado_hcyase.		

Db	277	IleCysGlyTyrGlyAspValGlyLysGlyCysAlaGluAlaAlaLysGlyGlnGlyAla	296
QY	871	CGCGTTATCATCACAGAACTCGACCAATCTCGCTCTCCAGGCTGCCATGAAGGCTAC	930
Db	297	ArgValThrIleThrGluIleAspProIleAsnAlaLeuGlnAlaLeuMetGluGlyPhe	316
QY	931	CAGTCCGCGCGCATCGAGGAAGTCGTCAGAGGATGTCGATATCTTCGTTCATGCGCAGGA	990
Db	317	AspValIleArgValGluAspValIleAlaAspSerAspIleValThrAlaThrGly	336
QY	991	AAGTCGGCATCATCTCTGTTTGACATGATGGCCCGCAGATGAAGATTAAGCGTATTTCGCT	1050
Db	337	AsnLysAspIleIleLeuLeuGluHisMetLysAlaMetLysAspHisAlaIleLeuGly	356
QY	1051	AACATCGGCGCACTTCGATAACGAAATGTATACAGATGGCTCATGAATAACCCAGGCATC	1110
Db	357	AsnIleGlyHisPheAspAsnGluIleAspMetAlaAlaLeuGluArg--SerGlyAla	375
QY	1111	AAGCACATCCCCAATCAAGCCAGAAATACGACATGTGGGAATTCACAGAT--GGCCACGCT	1167
Db	376	ThrArgLeuAsnIleLysProGlnValAspLeuThrThrPheGlyAspSerGlyLysSer	395
QY	1168	ATCCTCCTCTTCTGTGAGGGCGCGCTTCTTAACCTTGGCTGCGCTACAGCTACCCATCT	1227
Db	396	IleIleValLeuSerGluGlyArgLeuLeuAsnLeuGlyAsnAlaThrGlyHisProSer	415
QY	1228	TTCGTTATGTCAATGTCATTCACAAACCAGACACTCGCTGAGCTGACCTTACGAAAG	1287
Db	416	PheValMetSerAsnSerPheAlaAsnGlnThrIleAlaGlnIleGluLeuThrLys	435
QY	1288	AGAGAAATCTCGAGAAGAAGGTTTACACACTTCGGAAGCATTCGATGAAGAAGTCCT	1347
Db	436	AsnAspAspTyrAspAsnGluValTyrArgLeuProLysHisLeuAspGluLysValAla	455
QY	1348	CGCTCCACTCGATCTCTCGATGCCAGTCCACCTTACAAAGCTTACACAGAAGCAGGCTGAC	1407
Db	456	ArgValHisValGluAlaLeuGlyGlyGlnLeuThrLysLeuThrLysAspGlnAlaGlu	475
QY	1408	TATCATCAGGTTCCAGTTGAGGGTCTCTTACAGTCTGATGCTTACCGTTAT	1458
Db	476	TyrIleuGlyValAspValAspGlyProPheLysProAspHisTyrArgTyr	492
RESULT 8			
ID	SAHH ARATH	STANDARD;	PRT; 485 AA.
AC	O23255; O81847;		
DT	15-DEC-1998 (Rel. 37, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Adenosylhomocysteinease (EC 3.3.1.1) (S-adenosyl-L-homocysteine hydrolase) (AdoHcyase).		
DE	SAHH OR A74G13940 OR DL3010W.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=cv. Landsberg erecta;		
RC	Belbahr L., Elleuch H., Villarroel R., Inze D., Thomas D.,		
RC	Thomasset B.;		
FT	"The isolation of an Arabidopsis thaliana cDNA clone encoding S-		
RT	adenosyl-L-homocysteine hydrolase."		
RL	(in) Plant Gene Register PGR99-139.		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=cv. Columbia;		
RC	MEDLINE=98121113; PubMed=9461215;		
RA	Bevan M., Bancroft I., Bent E., Love K., Goodman H.M., Dean C.,		
RA	Bergkamp R., Dirks W., van Staveren M., Stiekema W., Drost L.,		
RA	Ridley P., Hudson S.-A., Patel K., Murphy G., Piffanelli P.,		
RA	Wedler H., Wedler E., Wambutt R., Weitzenegger T., Pohl T., Terryn N.,		

RA Gielen J., Villarroel R., De Clercq R., van Montagu M., Lechary A.,  
 RA Aubourg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S.,  
 RA Kottler P., Entian K.-D., Rieger M., Schaefer M., Funk B.,  
 RA Mueller-Auer S., Silvey M., James R., Monfort A., Pons A.,  
 RA Puigdomenech P., Douka A., Voukelatou E., Milioni D., Hatzopoulos P.,  
 RA Piravandi E., Obermaier B., Hilbert H., Dueterhoeft A., Moores T.,  
 RA Jones J.D.G., Eneva T., Palme K., Benes V., Rechmann S., Ansorge W.,  
 RA Cooke R., Berger C., Delseny M., Voet M., Volckaert G., Mewes H.-W.,  
 RA Klosterman S., Schueller C., Chalwatzis N.,  
 RA "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of  
 RT Arabidopsis thaliana";  
 RL Nature 391:485-488(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=20083488; PubMed=10617198;  
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,  
 RA Pohl T., Dueterhoeft A., Stiekema W., Entian K.-D., Terryn N.,  
 RA Harris B., Ansorge W., Brandt P., Grivell L.A., Rieger M.,  
 RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,  
 RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,  
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,  
 RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,  
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,  
 RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,  
 RA Braeken M., Wiltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,  
 RA Weitzenecker T., Boche G., Rampsberger U., Hilbert H., Braun M.,  
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirke W.,  
 RA Moollman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,  
 RA Bernaiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,  
 RA De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,  
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,  
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLeay K., Mayes R.,  
 RA Pettett A., Rajadream M.A., Lyne M., Benes V., Rechmann S.,  
 RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,  
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fartmann B., Grandeth K., Dauner D., Herzl A.,  
 RA Neumann S., Argiou A., Vitale D., Liguori R., Piravandi E.,  
 RA Massenet O., Quigley F., Clabaud G., Muendlein A., Feiber R.,  
 RA Schnabl S., Haller R., Schmidt W., Lechary A., Aubourg S.,  
 RA Chefor F., Cooke R., Berger C., Monfort A., Casacuberta E.,  
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,  
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacón D., Jesse T.,  
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,  
 RA Frishman D., Haase D., Lencke K., Mewes H.-W., Stocker S.,  
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,  
 RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,  
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,  
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
 RA Kramer J., Fulton L., Nardis E., Dente M., Pepin K., Hallier L.,  
 RA Neilson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,  
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,  
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,  
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,  
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Fill S.,  
 RA Granat S., Shondy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,  
 RA Chen E., Maria M., Martienssen R., McCombie W.R.;  
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
 RT thaliana";  
 RL Nature 402:769-777(1999).  
 RN [4]  
 RP SEQUENCE OF 19-485 FROM N.A.  
 RC STRAIN=cv. Landsberg erecta;  
 RA Zhang H., Forde B.G.;  
 RT "Identification of novel nitrate-inducible genes from Arabidopsis";  
 RL Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.  
 CC -!- FUNCTION: ADENOSYLHOMOCYSTEINE IS A COMPETITIVE INHIBITOR OF  
 CC S-ADENOSYL-L-METHIONINE-DEPENDENT METHYL TRANSFERASE REACTIONS;  
 CC THEREFORE ADENOSYLHOMOCYSTEINASE MAY PLAY A KEY ROLE IN THE  
 CC CONTROL OF METHYLATIONS VIA REGULATION OF THE INTRACELLULAR  
 CC CONCENTRATION OF ADENOSYLHOMOCYSTEINE (BY SIMILARITY).

CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine  
 CC + L-homocysteine.  
 CC -!- COFACTOR: NAD.  
 CC -!- PATHWAY: Activated methyl cycle.  
 CC -!- SUBUNIT: Homotrimer (By similarity).  
 CC -!- SIMILARITY: Belongs to the adenosylhomocysteinase family.  
 CC -----  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z97335; CAB10173.1; -.  
 CC EMBL; AF059581; AAC14714.1; -.  
 CC EMBL; AL161537; CAB78436.1; -.  
 CC EMBL; Z97059; CAB09795.1; -.  
 CC PIR; C71400; C71400.  
 CC HSP; P10760; 1B3R.  
 CC InterPro; IPR000043; Ado\_hcyase.  
 CC Pfam; PF05221; AdoHcyase; 1.  
 CC TIGRfam; TIGR00936; ahcY; 1.  
 CC PROSITE; PS00738; AdoHcyase\_1; 1.  
 CC PROSITE; PS00739; AdoHcyase\_2; 1.  
 CC Hydrolase; NAD; One-carbon metabolism.  
 CC NP\_BIND 263 294 NAD (POTENTIAL).  
 CC FT CONFLICT 80 84 E -> Q (IN REF. 4).  
 CC FT CONFLICT 96 96 A -> R (IN REF. 4).  
 CC FT CONFLICT 392 392 E -> Q (IN REF. 4).  
 CC FT CONFLICT 460 460 T -> R (IN REF. 4).  
 CC SQ SEQUENCE 485 AA; 53378 MW; 1113270A0F46C86C CRC64;  
 Alignment Scores:  
 Pred. No.: 1,86e-94 Length: 485  
 Score: 1537.00 Matches: 302  
 Percent Similarity: 74.33% Conservative: 60  
 Best Local Similarity: 62.01% Mismatches: 109  
 Query Match: 56.16% Indels: 16  
 DB: 1 Gaps: 4  
 US-09-759-990A-1 (1-1461) x SAHH\_ARATH (1-485)  
 Qy 19 ACTGTGCTCCATTCGAGTACAGAAATTCGAGACATCAACCTCCATGTTCTTCGGCGGTAAG 78  
 Db 8 ThrSerSerGlyArgGlyArgGlyValLysAspMetSerGlnAlaAspPheGlyArgLeu 27  
 Qy 79 GAACCTACCTTCTGAGAGAGAAATGCCAGCTTATGTTCTTCGTGAGCGGTATTC 138  
 Db 28 GluLeuGluLeuAlaGluValGluMetProGlyLeuMetAlaCysArgThrGluPheGly 47  
 Qy 139 GCTTCTAAGCCATTAAGGCTTCGAGATCTCTGTTCTCCCTCCATCAGATCCAGACA 198  
 Db 48 ProSerGlnProPheLysGlyAlaArgIleThrGlySerLeuHisMetThrIleGlnThr 67  
 Qy 199 GCGGTCTTCATCGAGACATCTCAGCTCTTGGTGTGTGATGTCAGATGGGCTTCTCGAAC 258  
 Db 68 AlaValLeuIleGluThrLeuThrAlaLeuGlyAlaGluValArgTrpCysSerCysAsn 87  
 Qy 259 ATCTTCTCTACACAGATACAGCGCTGCTGCTATCGTTGTCGGCCCAACAGGCACCA 318  
 Db 88 IlePheSerThrGlnAspHisAlaAlaAlaAlaAlaAlaArgAspSerAlaAla- ---- 105  
 Qy 319 GAGAAGCCAGCGGTATCCAGTCTTCCTGGAAGGGGCAACACTCCAGATACTGG 378  
 Db 106 -----ValPheAlaTrpLysGlyGluThrLeuGlnGlyTrp 118  
 Qy 379 GAGAACACATACCGCTCTCTCATGTCGCGCAGATGGTCAAGGCCCAACAGCGTTCGAT 438  
 Db 119 TrpCysThrGluArgAlaLeuAspTrpGlyProGlyGlyGlyProAspLeuIleValAsp 138

```

QY 439 GATGGTGGTATGCTACATCTCTCATCTCCAAAGGGCTTCGAA-----TTCCGAA 486
DQ 139 AspGlyGlyAspAlaThrLeuLeuLeuHisGluGlyValLysAlaGluGluLeuPheGlu 158
QY 487 ACAGCGGGTGTTCAGAGCAACAGAGCTGACACCTCCAAACCGCTGCGTCTTT 546
DQ 159 LysThrGlyGlnValProAspProThrSerThrAspSerProGluPheGlnLeuVal 178
QY 547 GCTACACTCAAGAGGCTTCCAAACCAAGACCAAGACCACTGGCACACAGTCTCCCGC 606
DQ 179 SerLeuLeuLysGluGlyLeuGlnValAspProLysLysTyrHisLysMetLysGlu 198
QY 607 ATGAACGGTGTTCGAGAGCAACAAACAGGTGTCACCGCTCTACACAGTTCGAGAG 666
DQ 199 LeuValGlyValSerGluGluThrThrThrGlyValLysArgLeuTyrGlnMetGln 218
QY 667 GAGGGCAACATCTCTTCCAGCCATCAACGTCACAGCGCTTCAAGTCCCAAGTTC 726
DQ 219 AsnGlyThrLeuLeuPheProAlaLeuAsnValAspSerValThrLysSerLysPhe 238
QY 727 GATAACATCTACGGCTGCGCCACCTCCCTTATCGATGATATCAACCGCTCTCCGATCT 786
DQ 239 AspAsnLeuTyrGlyCysArgHisSerLeuProAspGlyLeuMetArgAlaThrAsp 258
QY 787 ATGATCGGGGCAAGACAGCTCTCGTTCAGCGCATGTCGCGAAGGCTCGCCT 846
DQ 259 MetLeuAlaGlyLysValAlaValLeuLysGlyTyrGlyAspValGlyLysGlyCys 278
QY 847 CAATCCCTCCGGCGCAAGCGCTCGCGTTATCATCACAGCACTCCGACCAATCTGCG 906
DQ 279 AlaAlaMetLysThrAlaGlyAlaArgValLeuValThrGluLeuAspProLysCys 298
QY 907 CTCAGCGTCCATGAGAGCTACAGGTCGCGCCGATCGAGGAGTGTGCAAGATGTC 966
DQ 299 LeuGlnAlaLeuMetGluGlyLeuGlnValLeuThrLeuGluAspValValSerGlu 318
QY 967 GATATCTCGTTACATGACAGAACTCGGATATCATCTCTGTCATGATGCGCCAG 1026
DQ 319 AspLeuPheValThrThrThrGlyAsnLysAspLeuLeuMetValAspHisMetArg 338
QY 1027 ATGAAGTAAGGCTATTGTCGTAACATCGGCGCCACTTCGATACCAAGAAATTGATCAG 1086
DQ 339 MetLysAsnAsnAlaValCysAsnLeuGlyHisPheAspAsnGluLeuAspMetLeu 358
QY 1087 GGCTCATGAATACCGAGCATCAGCACATCCCAATCAAGCCAGACAGATACAGATG 1146
DQ 359 GlyLeuGluThrTyrProGlyValLysArgLeuThrLeuLysProGlnThrAspArg 378
QY 1147 GAATTCACAGATGCGCAGCT--ATCCTCTCTTGTGTCAGGCGCCCTTCTTAACCTT 1203
DQ 379 ValPheProGluThrLysAlaGlyLeuLeuValLeuAlaGluGlyArgLeuMetAsn 398
QY 1204 GGCTGCGTACAGGTACCCATCTTGTGTATGTCAATGTCATTCACAAACAGACATC 1263
DQ 399 GlyCysAlaThrGlyHisProSerPheValMetSerCysSerPheThrAsnGlnVal 418
QY 1264 GCTCAGCTGACCTCTACGAAAGACA-----GGAAATCTCGAGAGAGGTTTACACA 1317
DQ 419 AlaGlnLeuGluLeuThrPheAsnGluLysAlaSerGlyLysTyrGluLysValTyr 438
QY 1318 CTCGCAAGCATCTCGATGAAGAGTCTGCGCTCCACCTCGGATCTCTCGATGTCAC 1377
DQ 439 LeuProLysHisLeuAspGluLysValAlaLeuLeuHisLeuGlyLysLeuGlyAla 458
QY 1378 CTTACAAAGCTTACAGACAGCGTGAATCATCATCAGTTCGAGTTCAGGCTCTTAC 1437
DQ 459 LeuThrLysLeuSerLysAspGlnSerAspTyrValSerLeuProLysGluGlyPro 478
QY 1438 AAGTCTGATGCTTACCGTTAT 1458
DQ 479 LysProHisTyrArgTyr 485

```

RESULT 9

```

SAHH TOBAC
ID -SAHH TOBAC STANDARD; PRT; 485 AA.
AC P50248;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Adenosylhomocysteinease (EC 3.3.1.1) (S-adenosyl-L-homocysteine
DE hydrolase) (AdoHcyase) (Cytokinin binding protein CBP57).
GN SAHH.
OS Nicotiana tabacum (Common tobacco), and
OS Nicotiana sylvestris (Wood tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.
NCBI_TaxID=4097, 4096;
RN [1]
RP SEQUENCE FROM N.A.
RP SPECIES=N.tobacum; STRAIN=cv. Bright Yellow 4;
RA Tanaka H., Masuta C., Kataoka J., Kuwata S., Koiwai A., Noma M.;
RT "Inducible expression by plant hormones of S-adenosyl-homocysteine
RT hydrolase gene from Nicotiana tabacum during early flower bud
RT formation in vitro.";
RL Plant Sci. 113:167-174 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RP SPECIES=N.sylvestris;
RA Mitsui S., Wakasugi T., Sugiura M.;
RT "A cDNA encoding the 57 kDa subunit of a cytokinin-binding protein
RT complex from tobacco: the subunit has high homology to S-adenosyl-L-
RT homocystein hydrolase.";
RL Plant Cell Physiol. 34:1089-1096 (1993).
CC -!- FUNCTION: ADENOSYLOHOMOCYSTEINE IS A COMPETITIVE INHIBITOR OF
CC S-ADENOSYL-L-METHIONINE-DEPENDENT METHYL TRANSFERASE REACTIONS;
CC THEREFORE ADENOSYLOHOMOCYSTEINASE MAY PLAY A KEY ROLE IN THE
CC CONTROL OF METHYLATIONS VIA REGULATION OF THE INTRACELLULAR
CC CONCENTRATION OF ADENOSYLOHOMOCYSTEINE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine
CC + L-homocysteine.
CC -!- COFACTOR: NAD (BY SIMILARITY).
CC -!- PATHWAY: Activated methyl cycle.
CC -!- SIMILARITY: Belongs to the adenosylhomocysteine family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D45204; BAA08142.1; -
DR EMBL; D49804; BAA23164.1; -
DR EMBL; D16138; BAA03709.1; -
DR HSP; P10760; IBSR.
DR InterPro; IPR000043; Ado_hcyase.
DR Pfam; PF05221; AdoHcyase; 1.
DR Pfam; PF00670; AdoHcyase_NAD; 1.
DR TIGRfam; TIGR00936; ahcy_1.
DR PROSITE; PS00738; ADHOCYASE_1; 1.
DR PROSITE; PS00739; ADHOCYASE_2; 1.
KW Hydrolase; NAD; One-carbon metabolism.
FT NP BIND 263 294 NAD (POTENTIAL).
SQ SEQUENCE 485 AA; 53104 MW; AA6D6944E9DF0A5C CRC64;

```

Alignment Scores:

```

Pred. No.: 3,43e-94 Length: 485
Score: 1533.00 Matches: 305
Percent Similarity: 74.07% Conservative: 52
Best Local Similarity: 63.28% Mismatches: 109
Query Match: 56.01% Indels: 16
DB: 1 Gaps: 4

```

US-09-759-990A-1 (1-1461) x SAHH\_TOBAC (1-485)





QY	847	CAATCCCTCCGTGCGCAAGCGCTCCGTTATCATCAAGAACTCGACCCATCTCGCGCT	906
Db	279	AlaAlaLeuLysGlnAlaGlyValArgValIleValThrGluIleAspProIleCysAla	298
QY	907	CTCAGCGCTGCCATCGAAGGCTACCAAGTCCCGCCGATCGAGGAAGTCTCGAAGATGTC	966
Db	299	LeuGlnAlaLeuMetGluGlyIleGlnIleLeuThrLeuGluAspValValSerGluAla	318
QY	967	GATATCTCTGTTACATGCAAGAGAACTCGGATATCATCTCTGTTGACATGATGCCCC	1026
Db	319	AspIlePheValThrThrGlyArgValAspIleIleMetValAspHisMetArgLys	338
QY	1027	ATGAAGGATAAGGCTATTGTCGGTTAATCATCGCCACTTCGATACAGAAATTCATACAGAT	1086
Db	339	MetLysAsnAsnAlaIleValCysAsnIleGlyHisPheAspSangluIleAspMetAsn	358
QY	1087	GGCTCATGAATAATCCAGGCGATCAAGACATCCCAATCAAGCCAGAAATACACATGTGG	1146
Db	359	GlyLeuGluThrTyProGlyValLysArgIleThrIleLysProGlnThrAspArgTrp	378
QY	1147	GAATCCCGAGATGGCCAC--GGTATCCCTCTTGTGTGAGGGCGGCTCTTAACCTT	1203
Db	379	ValPheProGluThrLysThrGlyIleIleValLeuAlaGluGlyArgLeuMetAsnLeu	398
QY	1204	GGCTGCGCTACAGGTCACCCATCTTTGTTATGTCAATGTCAATCAACAAACAGACACT	1263
Db	399	GlyCysAlaThrGlyHisProSerPheValMetSerCysSerPheThrAsnGlnValIle	418
QY	1264	GCTCAGCTCGACTCTACGAAAGAGA-----GGAATCTCGAGAGAGAGTTTACACA	1317
Db	419	AlaGlnLeuGluLeuThrPAsnGluLysAlaSerGlyLysTrpGluLysLysValTrVal	438
QY	1318	CTTCCGAAGCATCTCGATGAAGAAGTCGCTCGCCCTCCACTCGGATCTTCGATGCCAC	1377
Db	439	LeuProLysHisLeuAspGluLysValAlaAlaLeuHisLeuGlyLysLeuGlyAlaArg	458
QY	1378	CTTACAAAGCTTACACAGAACGCGTGCATCATCAACGTTCCAGTTGAGGGTCTTAC	1437
Db	459	LeuThrLysLeuThrLysSerGlnSerAspTrpIleSerIleProIleGluGlyProTrp	478
QY	1438	AAGTCGTGATGCTTACCGTTAT	1458
Db	479	LysLeuArgLeuTrpArgTrp	485
RESULT 11			
SAHH CATRO	ID	SAHH CATRO	STANDARD; PRT; 485 AA.
AC	DT	01-FEB-1994 (Rel. 28, Created)	
DT	DT	01-FEB-1994 (Rel. 28, Last sequence update)	
DT	DT	28-FEB-2003 (Rel. 41, Last annotation update)	
DE	DE	Adenosylhomocysteinease (EC 3.3.1.1) (S-adenosyl-L-homocysteine hydrolase) (AdoHcyase).	
GN	GN	SAHH.	
OS	OS	Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).	
OC	OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vinales;	
OC	OC	Asteridae; Lamids; Gentianales; Apocynaceae; Rauvolfioideae; Vincaceae;	
OC	OC	Catharanthus.	
NCBI_TaxID=4058;	NCBI_TaxID=4058;		
SEQUENCE FROM N.A.	MEDLINE=94218405; PubMed=8165255;		
Schroeder G., Walz A., Horze M., Schroeder J.;	"cDNA for S-adenosyl-L-homocysteine hydrolase from Catharanthus		
roseus.";	Plant Physiol. 104:1099-1100(1994).		
-I- FUNCTION: ADENOSYLHOMOCYSTEINE IS A COMPETITIVE INHIBITOR OF			
S-ADENOSYL-L-METHIONINE-DEPENDENT METHYL TRANSFERASE REACTIONS;			
THESE ADENOSYLHOMOCYSTEINE MAY PLAY A KEY ROLE IN THE			
CONTROL OF METHYLATIONS VIA REGULATION OF THE INTRACELLULAR			
CONCENTRATION OF ADENOSYLHOMOCYSTEINE.			
-I- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O =adenosine			



```

CC + L-homocysteine.
CC -I- COFACTOR: NAD.
CC -I- PATHWAY: Activated methyl cycle.
CC -I- SUBUNIT: Homotetramer (By similarity).
CC -I- INDUCTION: By stress.
CC -I- SIMILARITY: Belongs to the adenosylhomocysteinase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: Z26881; CAA81527.1; ..
CC PIR: S38379; S38379.
CC HSSP: P10760; 1B3P.
CC InterPro: IPR000043; Ado hcyase.
CC InterPro: IPR000205; NAD_binding.
CC Pfam: PF05221; AdoHcyase; 1.
CC Pfam: PF00670; AdoHcyase NAD; 1.
CC TIGRFAMs: TIGR00936; ahcy; 1.
CC PROSITE: PS00738; ADHOCYASE_1; 1.
CC PROSITE: PS00739; ADHOCYASE_2; 1.
CC Hydrolase; NAD; One-carbon metabolism.
KW NP BIND 263 294 NAD (POTENTIAL).
FT SEQUENCE 485 AA; 53233 MW; 50C0B99A9F66C51 CRC64;
SQ
Alignment Scores:
Pred. No.: 2,5e-93 Length: 485
Score: 1520.00 Matches: 304
Percent Similarity: 74.18% Conservative: 58
Best Local Similarity: 62.30% Mismatches: 108
Query Match: 55.54% Indels: 18
DB: 1 Gaps: 6
US-09-759-990A-1 (1-1461) x SAHH_CATRO (1-485)
QY 19 ACTGTTGCTCCATTCAGTACAGAAATGCCGACATCAACCTCCATGTTCTCGCGCGTAAG 78
DB 8 ThrSerSerGlyArgGluTyrLysValLysAspMetSerGlnAlaAspPheGlyArgLeu 27
QY 79 GAACCTTACCTTGTCTGAGAGGAATGCCAGTCTTATGTTCTCTGTCGAGCGTTATCC 138
DB 28 GluileGluLeuAlaGluValGluMetProGlyLeuMetSerCysargAlaGluPheGly 47
QY 139 GCTTCTAGCCATTGAGGGGTGTCAGAAATCTCTGTTCCCTCCACATGACATCCAGACA 198
DB 48 ProSerGlnProPheLysGlyAlaLysIleThrGlySerLeuHisMetThrIleGlnThr 67
QY 199 GCGTCTCTCAGACACATCAGACTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 258
DB 68 AlaValLeuileGluThrLeuThrAlaLeuGlyAlaLeuGlyAlaValArgTyrCysSerCysasn 87
QY 259 ATCTTCTCTACACAGATACAGCCGCTGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 318
DB 88 IlePheSerThrGlnGluHisAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 105
QY 319 GAGAACCCAGCCGATCCAGTCTTCCGCTGAGAGGCGAACACTCCAGAACTCCG 378
DB 106 -----ValPheAlaTrpLysGlyGluThrLeuGlnGluTyrTrp 118
QY 379 GAGAACACATACCGGCTCTCACATGG-----CCAGATGTCTCAAGGCCACAGCAGGTTC 435
DB 119 TrpCysThrGluArgAlaLeuAspTrpGlyProAspGly---GlyProAspLeuileVal 137
QY 436 GATGATGGTGTGATGCTACACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 483
DB 138 AspAspGlyGlyAspAlaThrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 157
QY 484 GAAACAGCGGTGCTGTCTCCAGACCCACAGACAGTCTGACACCTCGAATACCGTTCGCTT 543
DB 484 GAAACAGCGGTGCTGTCTCCAGACCCACAGACAGTCTGACACCTCGAATACCGTTCGCTT

```

```

DB 158 LysLysAsnGlyAlaLeuProAspProSerSerThrAspAsnAlaGluPheGlnIleVal 177
QY 544 CTTCGTACTACTCAGCAGGTCTTCAACCAAGACAAGAACCAACCAACCAACCAACCAACCAAC 603
DB 178 LeuThrIlelelelelelelelelelelelelelelelelelelelelelelelelelele 197
QY 604 GGCATGACAGGTGTTTCCGAGAGACACAAACAGGTGTCACCCGCTCTACACAGCTCGAG 663
DB 198 ArgLeuValGlyValSerGluGluThrThrThrGlyValLysArgLeuTyrGlnMetGln 217
QY 664 AAGGAGGGCAAACTCTCTCCAGCCATCAACGTCACGAGCTGTACAAAGTCCAG 723
DB 218 AlaAsnGlyThrLeuLeuPheProAlaIleAsnValAsnAspSerValThrLysSerLys 237
QY 724 TTCGATACATCTTACGGTTCGCGCACTCCCTTATCGATGTGTATCAACCGTGTCTCCGAT 783
DB 238 PheAspAsnLeuTyrGlyCysArgHisSerLeuProAspGlyLeuMetArgAlaThrAsp 257
QY 784 GTCATGATCGCGGCAAGACAGCTCTCGTCATGGGTAGCGCGATGTCCGCAAGGGCTGC 843
DB 258 ValMetIleAlaGlyLysValAlaValAlaGlyTyrGlyAspValGlyLysGlyCys 277
QY 844 GCTCAATCCCTCGTGGCGCAAGCGCTCGCGTTATCATCACAGAACTCCGACCACTGC 903
DB 278 AlaAlaLeuLysGlnAlaGlyAlaArgValIleValThrGluIleAspProIleCys 297
QY 904 GCTCTCAGGTCGCCATGGAAGGTACAGGTCCGCGCATCGAGGAAGTGTCAAGAGAT 963
DB 298 AlaLeuGlnAlaThrMetGluGlyLeuGlnValLeuThrLeuGluAspValValSerGlu 317
QY 964 GTCGATATCTTGTGTACATGACAGGAACCTGCGATATCATCTCTGTGTGACATGATGGCC 1023
DB 318 AlaAspIlePheValThrThrThrGlyAsnLysAspIlelelelelelelelelelelelele 337
QY 1034 CAGATGAAGGATAGGCTATTGTCGTAACATCGGCCACTTCGATTAACGAAATGTATACA 1083
DB 338 LysMetLysAsnAsnAlaIleValCysAsnIleGlyHisPheAspAsnGluIleAspMet 357
QY 1084 GATGGCTCATGNAATACCCAGGATCAACACATCCCAATCCAGCCAGATACGACATG 1143
DB 358 LeuGlyLeuGluThrTyrProGlyValLysArgIleThrIleLysProGlnThrAspArg 377
QY 1144 TGGGAATTCACAGATGGCGCAAGCT---ATCCTCTTCTTGTGAGGGCGCCTTCTTAAC 1200
DB 378 TrpValPheProAspThrAsnSerGlyIlelelelelelelelelelelelelelelelelele 397
QY 1201 CTTCGCTCGCTACAGGTCAACCATCTTTCGTTATGTCAATGTCAATTCACAAACAGACACA 1260
DB 398 LeuGlyCysAlaThrGlyHisProSerPheValMetSerCysSerPheThrAsnGlnVal 417
QY 1261 CTCGCTCAGCTCGACTCTTACGAAAGAGA-----GGAAATCTCGAAGAGAGGTTTAC 1314
DB 418 IleAlaGlnLeuGluLeuThrPaspGluArgLysThrGlyLysTyrGluLysLysValTyr 437
QY 1315 ACATTCGGAAGCATCTCGATGAAGAAGTCTCGCTCGCTCCACCTCGGATCTCTCGATGTC 1374
DB 438 ValLeuProLysHisLeuAspGluLysValAlaAlaLeuHisLeuGlyLysLeuGlyAla 457
QY 1375 CACCTTCAAGGCTTACACAGAGCGGCTGACTACATCAAGTTCAGTTCAGTTCAGTTCCT 1434
DB 458 LysLeuThrLysLeuThrLysAspGlnAlaAspIlelelelelelelelelelelelelelele 477
QY 1435 TACAAGTCTGATCTTACCGTTAT 1458
DB 478 TyrLysProAlaHisTyrArgTyr 485
RESULT 12
SAHH_MESCR
ID SAHH_MESCR STANDARD; PRT; 485 AA.
AC P93253;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

```



```

Qy 1378 CTTACAAAGCTTACACAGAGCGGCTGACTACATCAACGTTCCAGTTGAGGGTCCTTAC 1437
Db 459 LeuThrLysLeuSerLysAspGlnAlaAspTyrIleSerValProValGluGlyProTyr 478
Qy 1438 AAGTCTGATGCTTACCGTTAT 1458
Db 479 LysProAlaHisTyrArgTyr 485

RESULT 13
SAHH_XANAC
ID _SAHH_XANAC STANDARD; PRT; 480 AA.
AC QPFP84;
DT 13-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE ADENOSYL-HOMOCYSTEINASE (EC 3.3.1.1) (S-adenosyl-L-homocysteine
DE hydrolase) (AdoHCyase).
DN AHCY OR SAHH OR XAC0804.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Canavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Stetual J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463 (2002).
CC -|- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine
CC + L-homocysteine.
CC -|- COFACTOR: NAD (By similarity).
CC -|- PATHWAY: Activated methyl cycle.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -|- SIMILARITY: Belongs to the adenosylhomocysteinase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AB011711; AM35692.1; -.
CC DR HMAP; MF 00563; -.
CC DR InterPro; IPR000043; Ado_hcyase.
CC DR Pfam; PF05221; AdoHCyase; 1.
CC DR Pfam; PF00670; AdoHCyase_NAD; 1.
CC DR TIGRfams; TIGR00936; aacy_1.
CC DR PROSITE; PS00738; ADOHCYASE_1; 1.
CC DR PROSITE; PS00739; ADOHCYASE_2; 1.
CC DR Hydrolase; NAD; One-carbon metabolism; Complete proteome.
CC FT NP BIND 261 292 NAD (POTENTIAL).
CC SQ SEQUENCE 480 AA; 52799 MW; 5EBBC690BA6C36B8 CRC64;

Alignment Scores: 1.15e-92 Length: 480
Pred. No.:

```

```

Score: 1510.00 Matches: 307
Percent Similarity: 74.79% Conservativity: 52
Best Local Similarity: 63.96% Mismatches: 105
Query Match: 55.17% Indels: 16
DB: 1 Gaps: 6

US-09-759-990A-1 (1-1461) x SAHH_XANAC (1-480)

Qy 34 GAGTACAGAAATGCCGACATCAACCTCCATGTTCTCGCGCTAAGGAACCTTACCCTTGCT 93
Db 12 AspTyrLysIleAlaAspIleSerLeuAlaAspTyrGlyArgLysGluLeuAspIleAla 31
Qy 94 GAGAAGGAATGCCAGGCTCTTATGGTTCTCGTGAGCGGTATTCGGCTTCTAAGCCATGG 153
Db 32 GluHisGluMetProGlyLeuMetSerIleArgAlaGlnThrLysProLeu 51
Qy 154 AAGGGTCTCAGAAATCTCTGGTTCCCTCCATGACAGATCCAGAGCGGCTTCATCGAG 213
Db 52 LysAspValArgLleThrGlySerLeuHisMetThrIleGlnThrAlaValLeuIleGlu 71
Qy 214 ACATCCACAGCTCTTGGTGTCTGATGTGAGATGGGCTTCTCGCAATCTTCTTACACAA 273
Db 72 ThrLeuLysAspIleGlyAlaAspValArgTyrAlaSerCysAsnIlePheSerThrGln 91
Qy 274 GATACAGCGCTGCTGCTATCGTTGTCGCCCAACAGGACACAGGACAGCCAGCGGT 333
Db 92 AspHisAlaAlaAlaAlaAlaAla-----ThrGlyThr----- 103
Qy 334 ATCCCGCTCTTGGCTCGAAGCGGCAACACTCCAGAAATCTGGAGAACACATACCCG 393
Db 104 ---ProValPheAlaThrLysGlyGluThrLeuGluGluTyrTyrAspCysThrLeuAsp 122
Qy 394 GCTCTCAATGG-----CGAGATGGT-----CAAGGCCACAGAGGTTGTCGATGAT 441
Db 123 AlaLeuThrPheThrLeuProAspGlyThrLeuThrGlyProGluLeuAlaValAspAsp 142
Qy 442 GGTGGTGTGCTACACTCTCTCATCTCCAGGGCTTCGAATTCGAAATCGAAACAGCGGTGCT--- 498
Db 143 GlyGlyAspValThrLeuLeuIleHisLysGlyTyrGluLeuGluAsnGlySerThrTyr 162
Qy 499 GTTCCAGAGCCAAACAGAGCTGACACACTCGAATACCGCTCGCTTCTTGTACATCTCAAG 558
Db 163 ValAspGluPro-----AlaSerSerHisGluGluGlyValIleLysAlaLeuLeuLys 180
Qy 559 CAGGTCTTCAACCAAGACAGAACCACTGGCACAGATGCTGCGCGCATGAGCGGTGT 618
Db 181 ArgValAlaValGluArgProGlyTyrTyrGlyValValValLysAspTyrLysGlyVal 200
Qy 619 TCCGAAGAGACAAACAGAGGTGTCCACCGCTCTACAGCTCGAGAGAGGAGGCAACTC 678
Db 201 SerGluGluThrThrThrGlyValHisArgLeuTyrGlnIleAlaGluAlaGlyLysLeu 220
Qy 679 CTCTTCCAGCCATCAAGTCAACAGCGCTGTGTACAAAGTCAAGTTTCGATATCATCTAC 738
Db 221 LeuIleProAlaIleAsnValAsnAspSerValThrLysSerLysPheAspAsnLeuTyr 240
Qy 739 GGTCTGCGCGCACTCCCTTATCGATGTGATCAACCGGTCTTCGATGTCATGATCGCGGC 798
Db 241 GlyCysArgGluSerLeuAlaAspGlyLeuLysArgAlaMetAspValMetLeuAlaGly 260
Qy 799 AAGACAGCTCTCTGTCATCGGTTACGCGGATGTGGCAAGGCTCGCTCAATCCCTCGGT 858
Db 261 LysValAlaValValCysGlyTyrGlyAspValGlyLysGlySerAlaAlaSerLeuArg 280
Qy 859 GGCAAGCGCTCGGTTATCATCAGAACTCGACCCATCTGGCTCTCCAGGGTGGC 918
Db 281 AlaTyrGlyAlaArgValIleValThrGluIleAspProIleCysAlaLeuGlnAlaSer 300
Qy 919 ATGGAAGGCTTACCAGGTCCGCGCATCGAGGAAGTGTGTCAAGGATGTCGATATCTTCGT 978
Db 301 MetGluGlyPheGluValAsnThrIleGluSerThrLeuGlyArgAlaAspIleTyrVal 320
Qy 979 ACATGCAGGAACACTGCGATATCATCTCTGTTGACATGATGGCCCGAGATGAAGGATAAG 1038

```

Db 321 ThrThrThrGlyAsnLysAspIleIleThrValGluHisLeuGlnAlaMetLysAspGln 340  
Qy 1039 GCTATTTCGGTAACATCGGCACCTCGATACGAAATTCATACAGATGGCTCATGAAA 1098  
Db 341 AlaIleValCysAenIleGlyHisPheAspAenGluIleGlnValAspAlaLeuLysAla 360  
Qy 1099 TACCCAGGATCAACAGCATCCCAATCAAGCCAGAAATACGACATGTGGGAATCCAGAT 1158  
Db 361 LeuLysAspValGlnLysIleAenIleLysProGlnValAspLysTyrValPheProAsn 380  
Qy 1159 GGCACCGATCCCTCTTCGTGAGGCGCCCTCTTAACTTCGCTGCGCTACAGGT 1218  
Db 381 GlyAsnAlaIlePheLeuLeuAlaAspGlyArgLeuValAsnLeuGlyCysAlaThrGly 400  
Qy 1219 CACCCATCTTCGTTAGTCAATGTCATTCACAAACAGACATCGCTCAGCTGACCTC 1278  
Db 401 HisProSerPheValMetSerAenSerPheAlaIleGlnThrLeuAlaGlnIleAspLeu 420  
Qy 1279 TACGAAAGAGAGGAAATCTCGAGAAAGGTTTACACATTCGGAAGCATCTCGATGAA 1338  
Db 421 TrpGluLysArgAspThrTyrGluLysLysValTyrIleLeuProLysHisLeuAspGlu 440  
Qy 1339 GAAGTCGCTCCCTCCCTCGATCTCGATCTCCATTCACAAAGCTTACAGAGTACAGAG 1398  
Db 441 GluValAlaArgLeuHisLeuGluLysIleGlyValLysLeuThrThrLeuThrLysAsp 460  
Qy 1399 CAGGCTGATCATCAACAGCTTCAGGTTGAGGCTCTTACAAAGTCTGATCTTACCGTAT 1458  
Db 461 GlnAlaAspTyrLeuGlyValAspValAlaGlyProLysLysProAspHisTyrArgTyr 480  
RESULT 14  
SAHH PHASS STANDARD; PRT; 485 AA.  
AC P50249;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Adenosylhomocysteinase (EC 3.3.1.1) (S-adenosyl-L-homocysteine  
hydrolyase) (AdoHcyase).  
GN SAHH.  
OS Phalaenopsis sp.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Orchidaceae;  
OC Epidendroideae; higher Epidendroideae; Vandaceae; Peridinae;  
OC Phalaenopsis.  
OX NCBI\_TaxID=36900;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95177653; PubMed=7872785;  
RA Preisig-Mueller R., Gnan P., Kindl H.;  
RT "The inducible 9, 10-dihydrophenanthrene pathway: characterization  
and expression of bibenzyl synthase and S-adenosylhomocysteine  
hydrolyase."  
RT Arch. Biochem. Biophys. 317:201-207(1995).  
RL Arch.  
CC -!- FUNCTION: ADENOSYLHOMOCYSTEINE IS A COMPETITIVE INHIBITOR OF  
S-ADENOSYL-L-METHIONINE-DEPENDENT METHYL TRANSFERASE REACTIONS;  
CC THEREFORE ADENOSYLHOMOCYSTEINASE MAY PLAY A KEY ROLE IN THE  
CC CONTROL OF METHYLATIONS VIA REGULATION OF THE INTRACELLULAR  
CC CONCENTRATION OF ADENOSYLHOMOCYSTEINE (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine  
+ L-homocysteine.  
CC -!- COFACTOR: NAD (BY SIMILARITY).  
CC -!- PATHWAY: Activated methyl cycle.  
CC -!- INDUCTION: BY INFECTION WITH B.CINEREA.  
CC -!- SIMILARITY: Belongs to the adenosylhomocysteinase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC  
CC  
DR EMBL; X79505; CAA56278.1; --  
DR PIR; S71621; S71621.  
DR HSSP; F10760; J63R.  
DR InterPro; IPR000043; Ado\_hcyase.  
DR Pfam; PF05221; AdoHcyase; 1.  
DR TIGRFAMs; TIGR00936; abcY; 1.  
DR PROSITE; PS00738; ADHOCYASE\_1; 1.  
DR PROSITE; PS00739; ADHOCYASE\_2; 1.  
KW Hydrolyase; NAD; One-carbon metabolism.  
FT NP\_BIND 263 294 NAD (POTENTIAL).  
SQ SEQUENCE 485 AA; 53141 MW; 1EB2CA5AD63AF233 CRC64;  
Alignment Scores:  
Pred. No.: 134e-92 Length: 485  
Score: 1509.00 Matches: 302  
Percent Similarity: 73.44% Conservative: 52  
Best Local Similarity: 62.66% Mismatches: 112  
Query Match: 55.13% Indels: 16  
DB: 1 Gaps: 4  
US-09-759-990a-1 (1-1461) x SAHH\_PHASS (1-485)  
Qy 34 GAGTACAGAAATGCCGACATCAACCTCATGTTCTCGGCGGTAGGAACCTTACCTTGCT 93  
Db 13 GluTyrLysValLysAspLeuSerGlnAlaAspPheGlyArgLeuGluIleGluLeuAla 32  
Qy 94 GAGAAAGAAATGCCAGGCTTATGTTCTTCGTGAGCGTTATTCGCTTCTAAAGCCATTG 153  
Db 33 GluValGluMetProGlyLeuValAlaCysArgAlaGluPheGlyProSerGlnProPhe 52  
Qy 154 AAGGTCATCAATCTCTGTTCCCTCCATGACATGACAGTCCAGACAGCGCTCTCATCGAG 213  
Db 53 LysGlyAlaArgIleSerGlySerLeuHisMetThrIleGlnThrAlaValLeuLeuGlu 72  
Qy 214 ACATCTACAGCTCTGCTGCTGATGTCAGATGGCTTCTCGACATCTTCTTACACAA 273  
Db 73 ThrLeuThrAlaLeuGlyAlaGluValArgTrpCysSerCysAsnIlePheSerThrGln 92  
Qy 274 GATACAGCCCTGCTCTATGTTGTCGGCCCAACAGGACACAGGAGGAGCCAGCGGT 333  
Db 93 AsphIsalaAlaAlaAlaAlaArgAspSerAlaAla----- 105  
Qy 334 ATCCAGCTCTTCGCTGAGGCGGAAACHCTCCAGATCTGGAGAACACATACCCG 393  
Db 106 -----ValPheAlaTrpLysGlyGluThrLeuGlnGluTyrTrpTrpCysThrGluArg 123  
Qy 394 GCTCTCATGCGCAGATGTCGAAGGCCACAGCAGGTTGTCGATGATGTTGATGCT 453  
Db 124 CysLeuGluTrpGlyAlaGlyGlyProAspLeuIleValAspAspGlyGlyAspAla 143  
Qy 454 ACATCTCTCATCTCCAGGGCTTC-----GAATTCGAAACAGCGGCTGCTGT 501  
Db 144 ThrLeuLeuIleHisGluGlyValLysAlaGluGluTyrGluLysAsnGlyLysIle 163  
Qy 502 CCAGAGCCACACAGATCGAACCTCGAATACCGCTGCGGTCTTCTTACACTCAAGCAG 561  
Db 164 ProAspProAlaSerThrAspAsnAlaGluPheGlnIleValLeuGlyLeuIleArgAsp 183  
Qy 562 GTCTTCAACCAACAGAACCCACTGGCCACACAGTTCCTGCGCGCATGACCGGTGTTCC 621  
Db 184 SerLeuSerValAspProLysLysTyrArgArgWetLysGluValGluValSer 203  
Qy 622 GAGAGACACACAGAGTGTCCACGGCTCTACAGCTCGAGAGGAGGCAACTCCTC 681  
Db 204 GluGluThrThrGlyValLysArgLeuTyrGlnMetGlnTyrSerGlyThrLeuLeu 223  
Qy 682 TTCCCGCCATCAAGCTCAACGAGCTGTTACAAAGTCCAGTTCGATTAACATCTACGC 741  
Db 224 PheProAlaIleAsnValAsnAspSerValThrLysSerLysPheAspAsnLeuTyrGly 243



```
QY 502 CCAGAGCCAAACAGAGCTGACAACTCGAATACCCGCTTCCTGCTACACTCAAGCAG 561
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 164 ProAspProAspSerThrAspAsnAlaGluPheIleValLeuSerIleIleIleGlu 183
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 562 GTCTTCAACCAAGACAGACCACTGGCAGCAGTGTGCTGCCGGCATGAACGGTGTTC 621
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 184 GlyLeuLysThrAspProLysArgTyrHisLysMetLysAspArgValValGlyValSer 203
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 622 GAAGACACAACAAGAGTGTCCACCGCTCTACAGCTCGAGAGGAGGAGGCAAACTCCTC 681
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 204 GluGluThrThrThrGlyValLysArgLeuTyrGlnMetGlnAlaAsnGlyThrLeuLeu 223
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 682 TTCGAGCCATCAACGTCACAGCGCTGTACAAAGTCCAAAGTTCGATACATCTACGGC 741
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 224 PheProAlaIleAsnValAsnSerValThrLysSerLysPheAspAsnLeuTyrGly 243
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 742 TGGCGCCACTCCCTTATCGATGTATCAACCGTGTCCGATGTCATGATCGCGGCAAG 801
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 244 CysArgHisSerLeuProAspGlyLeuMetArgAlaThrAspValMetIleAlaGlyLys 263
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 802 ACAGCTCTCGTATCGGTTACGGCGATGTGGCAAGGCTGCGCTCAATCCCTCGTGGC 861
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 264 ValAlaValAlaGlyTyrGlyAspValGlyLysGlyCysAlaAlaLeuLysGln 283
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 862 CAAGGGCGCTCGGTTATCATCAGACACTCGACCCAACTCCGCTCTCCAGGCTGCCATG 921
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 284 AlaGlyAlaArgValIleValThrGluIleAspProIleCysAlaLeuGlnAlaThrMet 303
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 922 GAAGGCTACAGGTCGCGCATCGAGGAAGTCGCAAGGATGCGATATCTTCGTACA 981
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 304 GluGlyLeuGlnValLeuThrLeuGluAspValValSerGluAlaAspIlePheValThr 323
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 982 TGCACAGGAAATGCGATATCATCTCTGTGACATGATGGCCACAGATGAAGGCT 1041
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 324 ThrThrGlyAsnLysAspIleIleMetLeuAspHisMetLysLysMetLysAsnAla 343
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1042 ATTGTGCGTAACATCGGCTCATGATACGAAATGATACAGATGCGCTCATGAATAC 1101
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 344 IleValCysAsnIleGlyHisPheAspAsnGluIleAspMetLeuGlyLeuGluThrHis 363
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1102 CCAGGCATCAAGCATCCCAATCAAGCCAGAAATACGACATGTGGGAATTCACAGATGC 1161
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 364 ProGlyValLysArgIleThrIleLysProGlnThrAspArgTrpValPheProGluThr 383
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1162 CAC---GCTATCCTCTTCGTCGAGGCGCGCTCTTAACCTTGGCTGCGCTACAGT 1218
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 384 AsnThrGlyIleIleIleLeuAlaGluGlyArgLeuMetAsnLeuGlyCysAlaThrGly 403
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1219 CACCCATCTTCGTTATGTCATTCATTCACAAACACAGACATCGCTCAGCTCGACCTC 1278
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 404 HisProSerPheValMetSerCysSerPheThrAsnGlnValIleAlaGlnLeuGluLeu 423
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1279 TAGAAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1332
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 424 TrpAsnGluLysSerSerGlyLysTyrGluLysLysValTyrValLeuProLysHisLeu 443
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1333 GATGAAGAGTCTCGCTCCACCTCGGATCTCTCGATGTCACCTTACAAAGCTTACA 1392
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 444 AspGluLysValAlaAlaLeuHisLeuGluLysLeuGlyAlaLysLeuThrLysLeuSer 463
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1393 CAGAAGCAGGCTGACTACATCAAGTTCAGTTGAGGCTCCTTACAAGTCTGATGCTTAC 1452
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 464 LysAspGlnAlaAspTyrIleSerValProValGluGlyProTyrLysProPheHisTyr 483
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1453 CGTTAT 1458
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 484 ArgTyr 485
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: October 15, 2003, 14:41:19 ; Search time 153 seconds  
(without alignments)  
4928.299 Million cell updates/sec

Title: US-09-759-990A-1

Perfect score: 2737

Sequence: 1 atggcttgcaatcacctac.....ctgatgcttaccgttattaa 1461

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delcp 6.0 , Delcxt 7.0

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 1661050

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O=/cgn2 1/USPTO.spool/US0975990/runat\_15102003\_115153\_2371/app\_query.fasta.1.1607  
-DB=SPTREML 23 -OFT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=rspt -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFT=ptp -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09759900.CGN.1.1.207@runat\_15102003\_115153\_2371 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREML 23.\*

1: sp.archaea.\*  
2: sp.bacteria.\*  
3: sp.fungi.\*  
4: sp.human.\*  
5: sp.invertebrate.\*  
6: sp.mammal.\*  
7: sp.mtc.\*  
8: sp.organelle.\*  
9: sp.phage.\*  
10: sp.plant.\*  
11: sp.rodent.\*  
12: sp.virus.\*  
13: sp.vertebrate.\*  
14: sp.unclassified.\*  
15: sp.rvirus.\*  
16: sp.bacteriap.\*  
17: sp.archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1580	57.7	469	2	Q8GGL7 streptomyces

2	1555	56.8	471	16	Q8KEG8
3	1536	56.1	485	10	Q8LE20
4	1512	55.2	485	10	Q8LPS8
5	1512	55.2	485	10	Q94KS
6	1512	55.2	485	10	Q9LK36
7	1510	55.2	480	16	Q8PPB4
8	1504	55.0	485	10	Q94929
9	1500	54.8	480	16	Q8PCH5
10	1483	54.2	478	16	Q8PRJ4
11	1475	53.9	450	10	Q42939
12	1454	53.1	474	16	Q8NSC4
13	1355.5	49.5	463	2	Q8KYV9
14	1351.5	49.4	436	16	Q8EXV1
15	1346	49.2	466	16	Q8FXZ7
16	1335	48.8	479	5	Q8MUG1
17	1279	46.7	432	6	Q8HXL1
18	1266.5	46.3	432	5	Q9VXV5
19	1253.5	45.8	432	5	Q8MZ11
20	1088.5	39.8	492	5	Q8MYX7
21	1057	38.6	521	5	Q9VZX9
22	1042	38.1	508	11	Q8BIH1
23	1032	37.7	530	4	Q96PK4
24	1032	37.7	597	4	Q9UG84
25	1016	37.1	431	10	Q947H3
26	906.5	33.1	324	11	Q8BF17
27	905.5	33.1	312	10	Q9SDP1
28	869.5	31.8	285	4	Q9H4U6
29	779	28.5	429	16	Q8DGC8
30	730.5	26.7	411	17	Q8PUQ4
31	704.5	25.7	411	17	Q8TEA5
32	703.5	25.6	425	2	Q8GDW5
33	676.5	24.7	195	10	Q9XE18
34	598	21.8	181	2	Q9R6R6
35	571.5	20.9	232	10	Q8W530
36	521	19.0	171	10	Q9SP98
37	518.5	18.9	218	4	Q43210
38	494	18.0	140	2	Q9R6S7
39	463.5	16.9	202	4	Q9BT10
40	413.5	15.1	500	16	Q8G5A1
41	402.5	14.7	143	4	Q9BZ13
42	386	14.1	121	10	Q41974
43	350	12.8	90	10	Q9M4V0
44	346	12.6	102	10	Q9XF45
45	322	11.8	96	2	Q93CC0

## ALIGNMENTS

RESULT 1  
Q8GGL7 PRELIMINARY; PRT; 469 AA.  
ID Q8GGL7  
AC Q8GGL7  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Adenosylhomocysteinase.  
OS Streptomyces atroolivaceus.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=66869;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22336326; PubMed=12446651;  
RA Cheng Y.O., Tang G.L., Shen B.;  
RT "Identification and Localization of the Gene Cluster Encoding  
RT Biosynthesis of the Antitumor Macrolactam Leinamycin in Streptomyces  
RT atroolivaceus S-140.";  
RL J. Bacteriol. 194:7013-7024 (2002).  
DR EMBL; AF484556; AAN85548.1; -.  
SQ SEQUENCE 469 AA; 51088 MW; 9B1980B1A5PCFA2E CRC64;

Alignment Scores:

Pred. No.: 1 1580 57.7 469 2 Q8GGL7 Length: 469

Score: 1580.00 Matches: 307  
 Percent Similarity: 77.26% Conservative: 60  
 Best Local Similarity: 64.63% Mismatches: 92  
 Query Match: 57.73% Indels: 16  
 DB: 2 Gaps: 3

US-09-759-990A-1 (1-1461) x Q8GGL7 (1-469)

```

QY 34 GAGTACAGAAATTCGCGACATCAACCTCCATGTTCTCGCGCGTAAGCAACTTACCTGGCT 93
Db 11 AspPheLysValAlaAlaAspLeuSerLeuAlaAlaPheGlyArgGlyLeuIleThrLeuAla 30
QY 94 GAGAAGAAATGCCAGAGCTTTATGTTCTTGTGAGGCTTATTCCTCCCTTCTAAGCAATG 153
Db 31 GluHisGluMetProGlyLeuMetSerIleArgGluGluTyrAlaAlaAlaGlnProLeu 50
QY 154 AAGGTGTCAAGAACTCTGTTCCCTCCACATGACATCCAGACAGCGCTCTCATCGAG 213
Db 51 AlaGlyAlaAlaGIIeThrGlySerLeuHisMetThrValGlnThrAlaValLeuIleGlu 70
QY 214 ACATCAAGAGCTCTTGGTGTGATGTCAGATGGGCTTCCTGCAACATCTTCTTACACAA 273
Db 71 ThrLeuValAlaLeuGlyAlaAspValArgTrpValSerCysAsnIleTyrSerThrGln 90
QY 274 GATACAGCGCTGTCTATCTGTTGTGCGCCACAGACGACACAGAGCCAGCCGGT 333
Db 91 AspHisAlaAlaAlaAlaIleAlaAla-----AlaGly 101
QY 334 ATCCAGCTCTTGGCTGGAAGGCGGAAACATCCAGAAATCTGGGAGAGACATATACCGC 393
Db 102 IleProValPheAlaTrpLysGlyGluThrLeuGluGluTyrTrpCysThrGluGln 121
QY 394 GCTCTCATAGCGCAGATGTCTAAGGCCACAGCAGAGTTGTCTGATGATGGTGTGATGCT 453
Db 122 AlaLeuThrTrpProGlyHisThrGlyProAsnMetIleLeuAspGlyGlyAla 141
QY 454 ACATCTCTCATCTCAAGGCTTCGAAATCGAAACAGCGGTGCTGTCAGAGCCACAA 513
Db 142 ThrLeuLeuValHisLysGlyValGluTyrArgLysThrGlyLeuPro----- 158
QY 514 GAAGCTGACAACTCGAATACCGCTGCTTCTGCTACTCACTCAAGCAGGCTTTCACACAA 573
Db 159 GluAlaGluAsnGlu-----LeuAlaValAlaArgAlaLeuLeuAspArg 174
QY 574 GACAGAACCATCGCACACAGTGTGTCGCGCATGACAGGTCTTCGAAAGAGACAA 633
Db 175 SerGlyLeuAspTrpThrAlaMetSerGluIleArgGlyValThrGluGluThr 194
QY 634 ACAGGTGTCCAGCGCTCTACAGCTCGAAGAGGAGGCAACTCTCTTCCAGCCATC 693
Db 195 ThrGlyValHisArgLeuTyrGluMetHisArgAspGlyThrLeuLeuPheProAlaIle 214
QY 694 AAGTCAAGCAGCTGTACAAAGTTCAGATTCAGATACATCTACGGCTGCGCGCACTCC 753
Db 215 AsnValAsnAspAlaValThrLysSerLysPheAspAsnLysTyrGlyCysArgHisSer 234
QY 754 CTTATCGATGTATCAACCGTGTCTCCGATGTCATGATCGCGCGCAAGACACTCTGTC 813
Db 235 LeuIleAspGlyIleAsnArgAlaThrAspValLeuIleGlyLysThrAlaValVal 254
QY 814 ATGGGTACCGCATGTGCGCAAGGCTCGCTCAATCCCTCCGTCGCAAGCGCTGCG 873
Db 255 CysGlyTyrGlyAspValGlyLysGlyCysAlaGluSerLeuArgGlyGlnGlyAlaArg 274
QY 874 GTTATCATCAGAACTCGACCAATCTCGCTCTCCAGGCTCGCATGGAAGCTTACAG 933
Db 275 ValIleIleThrGluIleAspProIleCysAlaLeuGlnAlaAlaMetAspGlyTyrGln 294
QY 934 GTCCGCGCATCGAAGAGTGTCAAGGATGTCTGATTCCTTTCATGTCACGACAGAAAC 993
Db 295 ValAlaThrLeuAspGluValAlaAspLysAlaAspIlePheValThrThrGlyAsn 314
QY 994 TGGCATATCATCTCTGTGTGATGATGTCGCCGAGATGAAGGATAGGCTATTGTCGGTAAC 1053

```

```

Db 315 LysAspIleIleMetAlaSerAspMetAlaLysMetLysHisGlnAlaIleValGlyAsn 334
QY 1054 ATCGGCCACTTCGATAAGAAATTCATACAGATGGCTTCATGAATACCCAGGCATCAG 1113
Db 335 IleGlyHisPheAspAsnGluIleAspMetAlaGlyLeuAlaGlnIleProGlyIleVal 354
QY 1114 CACATCCCAATCAAGCCAGAAATACGATGTGGATGTCATCCAGATGGCCACGCTATCCTC 1173
Db 355 LysAspGluValLysProGlnValHisThrTrpLysPheProAspGlyLysValLeuIle 374
QY 1174 CTTCTCTGAGGCGCGCTCTTAACTTGGCTGCTGCTACAGGTCAACCATCTTTCGTT 1233
Db 375 ValLeuSerGluGlyArgLysLeuAsnLeuGlyAsnAlaThrGlyHisProSerPheVal 394
QY 1234 ATGTCATATGTCATTCACAAACACACACTCGCTCAGCTCGACCTCTACGAAAGAGAGA 1293
Db 395 MetSerAsnSerPheAlaAspGlnThrLeuAlaGlnIleGluLeuPheThrLysProAsp 414
QY 1294 AATCTGAGAGAGAGGTTTACACACTTCCGAGCATCTCGATGAAGAGTCGCTCGCTC 1353
Db 415 GluTyrProThrAspValTyrValLeuProLysHisLeuAspGlyLysValAlaAspGly 434
QY 1354 CACCTCGCATCTCGATGTCACCTTACAAAGCTTACACAGAGCAGGCTGACTCATC 1413
Db 435 HisLeuAspAlaLeuGlyValLysLeuThrLeuArgProGluGlnAlaSerTyrIle 454
QY 1414 AACCTTCCAGTTCGAGGCTCTTCAAGTCTGATGCTTACCGCTTAT 1458
Db 455 GlyValGluValAspGlyProTyrLysProAspHisTyrArgTyr 469

```

RESULT 2

Q8KEG8 ID Q8KEG8 PRELIMINARY; PRT; 471 AA.

AC Q8KEG8; 01-OCT-2002 (TrEMBLrel. 22, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Adenosylhomocysteinase.

GN SAHH OR CT0721.

OS Chlorobium tepidum.

OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;

OC Chlorobium.

OX NCBI\_TaxID=1097;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=TLS / ATCC 49652 / DSM 12025;

RX MEDLINE=22103685; PubMed=12093901;

RA Eissen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,

RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Kolonay J.L., Yang F.,

RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Parksey D.,

RA Holt I., Unayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,

RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,

RA Vamathavan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,

RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;

RT "The complete genome sequence of Chlorobium tepidum TLS, a

RT photosynthetic, anaerobic, green-sulfur bacterium.";

EL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).

DR EMBL; AB012843; AM71958.1; -.

DR TIGR; CT0721; -.

DR InterPro; IPR000043; Ado hcyase.

DR InterPro; IPR000205; NAD binding.

DR Pfam; PF00670; AdoHcyase; 1.

DR TIGRfams; TIGR00936; ahcy; 1.

DR PROSITE; PS00738; ADHOCYASE\_1; 1.

DR PROSITE; PS00739; ADHOCYASE\_2; 1.

KW Complete proteome.

SQ SEQUENCE 471 AA; 51949 MW; 31E082405D241047 CRC64;

Alignment Scores:  
 Pred. No.: 1,01e-106 Length: 471  
 Score: 1555.00 Matches: 303  
 Percent similarity: 76.36% Conservative: 62



Best Local Similarity:	63.39%	Mismatches:	101
Query Match:	56.81%	Indels:	12
DB:	16	Gaps:	3
US-09-759-990A-1 (1-1461) x Q8KEG8 (1-471)			
Qy	25	GCTCCATTCGAGTACAGAAATTCGCGACATCAACCTCCATGTTCTCGCGCGTACGAACTT	84
Db	6	AlaValLeuAspTyrLysValAlaAspLysSerLeuAlaGluTrpGlyArgLysGluIle	25
Qy	85	ACCCCTGCTCAGAGAAGAAATGCCAGGCTTATGGTTCTTCGTGAGCGCTTATTCGCGCTCT	144
Db	26	GluIleAlaGluLysGluMetProGlyLeuMetAlaThrArgLysLysTyrGluGlyLys	45
Qy	145	AAGCCATTGAAGGCTCAGAACTCTCGTTTCCTCCACATCAGACTCCAGACGCCCTC	204
Db	46	LysProLeuAlaGlyAlaArgIleAlaGlySerLeuHisMetThrIleGlnThrAlaVal	65
Qy	205	CTCATCCAGACACTCACAGCTCTGGTGGCTGATGCAGATGGGTTCTCCTCAACATCTC	266
Db	66	LeuIleGlnThrLeuValGluLeuGlyAlaAspValArgTrpAlaSerCysAsnIlePhe	85
Qy	265	TCCTACACAGATCACCGCTCTCTCTATCGTTGTTCGGCCCAACAGGCAACCCAGAGAAG	324
Db	86	SerThrGlnAspHisAlaAlaAlaIleAlaAla-----	97
Qy	325	CCAGCCGGTATCCAGCTCTTCGCTGGAAGGGCGAAACACTCCAGAAATCTGGGAGAAC	384
Db	98	---AlaGlyValProValPheAlaTrpLysGlyGluThrLeuAspGluTrpTrpCys	116
Qy	385	ACATACCCGCGCTCTCATGTGGCCAGATGGTCAAGGCCCAACAGAGTTGTTCGATGATGT	444
Db	117	ThrArgGlnIleLeuGluPheGluGlyLeuGlyProAsnLeuIleValAspAspGly	136
Qy	445	GGTGATGCTCACTCTCTATCTCCAGGGCTTCGAATTCGAAACAGCGGTGCTGTTCCA	504
Db	137	GlyAspAlaThrLeuMetIleHisPheGlyTyrLysIleGluAsnAspProSerMetLeu	156
Qy	505	GAGCCAAACAGAGCTCAACCTCGAATACCGTGGGTTCTTGTACACTCAACAGACAGTCT	564
Db	157	AspLysThrProGly---AsnAlaGluLysAlaLeuLeuGlnLeuLysAlaVal	175
Qy	565	TTCAACACAGAGAAGAACCACTGGGACACAGTTGCTCGCGCGATGACAGGTGTTCCGAA	624
Db	176	PheAlaGluAspAsnGlnArgTrpHisLysValAlaAlaGlyMetLysGlyValSerGlu	195
Qy	625	GAGACAAACAGAGTCTCCAGCGCTTACACAGCTCGAGAAGAGGGCAAACTCCCTTC	684
Db	196	GluThrThrThrGlyValHisArgLeuTyrGlnMetMetGluLysGlyGluLeuLeuPhe	215
Qy	685	CCAGCCATCAAGCTCAACGAGCTGTATCAAAAGTCCAAAGTTCGATTAACATCTACGGTGC	744
Db	216	ProAlaIleAsnValAsnAspSerValThrLysSerLysPheAspAsnLeuTyrGlyCys	235
Qy	745	CGCCACTCCCTTATCATGATGATCAACCGGTCTTCGATGTCATGTCGCGCGGACAGACA	804
Db	236	ArgGluSerLeuAlaAspGlyIleLysArgAlaThrAspValMetIleAlaGlyLysVal	255
Qy	805	GCCTCTCTGATGGTTACGGCGATGTCGGCAAGGGCTGGGCTCAATCCCTCCGTGGCCAA	864
Db	256	ValValValLeuGlyTyrGlyAspValGlyLysGlyCysAlaHisSerMetArgSerTyr	275
Qy	865	GGCGCTCGGTATCATCAGAACTCGACCCCAATCTGGCTCTCCAGGTGCGATGGAA	924
Db	276	GlyAlaArgValIleValThrGluIleAspProIleCysAlaLeuGlnAlaAlaMetGlu	295
Qy	925	GGCTACCAAGTTCGCGCATCCAGGAAGTCTGATCAAGGATGTCGATATCTTCGTTACATGC	984
Db	296	GlyPheGluValThrThrMetGluAlaValLysGluGlyAsnIlePheValThrAla	315
Qy	985	ACAGAAACTCGGATATCATCTCTGTTCACATGATGCGCCAGATGAAGGATTAAGCTATT	1044
Db	316	ThrGlyAsnLysAspValIleThrLeuAspHisLysGlnMetArgAspGluAlaIle	335

QY	1045	GTGCGTATACATCGGCCACTTCGATACAGAAATTTGATACAGATGGCTCATGAATACCA	1101
Db	336	ValCysAsnIleGlyHisPheAspAsnGluIleGlnValAspAlaLeuAsnAsnPhelys	355
QY	1105	GGCATCAAGCATCCCAATCAAGCCAGAAATACGATGTGGGAATTCACGATGGCCAC	1164
Db	356	GlyAlaThrArgIleAsnIleLysProGlnValAspIleValPheGluAsnGlyAsn	375
QY	1165	GCTATCTCTCTTCTGCTGAGGGCGGCTCTTTAACTTCGTGGCTGCGCTACAGCTCACCCA	1224
Db	376	CysIleTyrLeuLeuAlaGluGlyArgLeuValAsnLeuGlyCysAlaThrGlyHisPro	395
QY	1225	TCCTTCGTTATGTCATCTCAATTCACAAACCAGACACTCGCTCAGCTCGACCTCTACGAA	1284
Db	396	SerPheValMetSerAsnSerPheThrAsnGlnThrLeuAlaGlnIleGluLeuTrpGln	415
QY	1285	AAGAGAGGAAATCTCGAGAAAGAGTTTACACACTTCGAGAGCTTCGATGAAGAATCT	1344
Db	416	-----AsnAspTyrLysValGlyValTyrArgLeuProLysLysLeuAspGluVal	433
QY	1345	GCTCGCTCCACTCGGATCTCTCGATGTCACCTTCCACCTTACAAGCTTACAGAGCAGCT	1404
Db	434	AlaArgLeuHisLeuGlyGlnIleGlyAlaLysLeuThrThrLeuThrLysGluGlnAla	453
QY	1405	GACTACATCAACAGTTCCCAAGTTGAGGGTCTCTTCAAGTCTGATGCTTACCGTTAT	1458
Db	454	AspTyrIleGlyValProValGluGlyProTyrLysProGluHisTyrIleArgTyr	471
RESULT 3			
ID	Q8LE20	PRELIMINARY; PRT; 485 AA.	
AC	Q8LE20;		
DT	01-OCT-2002 (TrEMBLrel. 22, Created)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE	Adenosylhomocysteinase.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.		
OX	NCBI_TaxID=3702;		
RP	[1]	SEQUENCE FROM N.A.	
RP	Haas B.J., Volkovsky N., Town C.D., Troukhan M., Alexandrov N.,		
RA	Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;		
RT	"Full-length messenger RNA sequences greatly improve genome		
RT	annotation."		
RL	Genome Biol. 0:0-0(2002).		
RP	[2]	SEQUENCE FROM N.A.	
RP	Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,		
RA	Feldmann K.;		
RT	"Full-length cDNA from Arabidopsis thaliana.;"		
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBSJ databases.		
DR	EMBL, AY085669; AAM62888.1; -.		
DR	InterPro; IPR000043; Ado_hcyase.		
DR	Pfam; PF00670; AdoHcyase; 1.		
DR	TIGRFAMS; TIGR00936; ahcy; 1.		
DR	PROSITE; PS00738; ADOHCYASE 1; 1.		
DR	PROSITE; PS00739; ADOHCYASE 2; 1.		
SQ	SEQUENCE 485 AA; 53401 MW; 11132B8C1F46C86C CFC64;		
Alignment Scores:			
Pred. No.:	2,626-105	Length:	485
Score:	1536.00	Matches:	302
Percent Similarity:	74.13%	Conservative:	59
Best Local Similarity:	62.01%	Mismatches:	110
Query Match:	56.12%	Indels:	16
DB:	10	Gaps:	4

US-09-759-990A-1 (1-1461) x Q8LE20 (1-485)

US-09-759-990A-1 (1-1461) x Q8LE20 (1-485)

QY 19 ACTGGTCTCCATCGAGTACAGATTGCCGACATCAACCTCCATGTTCTCGCGCGTAAG 78  
 Db 8 ThrSerSerGlyArgGluTyrLysValLysAspMetSerGlnAlaAspPheGlyArgLeu 27  
 QY 79 GAATTTACCTTCCTGAGAGGAAATGCCAGCTCTATGTTCTTCTGTCAGCGTATTCC 138  
 Db 28 GluLeuGluLeuAlaGluValGluMetProGlyLeuMetAlaCysArgThrGluPheGly 47  
 QY 139 GCTTTCAACCAATTGAGGGTGTCAAGATCTCTGGTTCCTCCACATGACAGTCCAGACA 198  
 Db 48 ProSerGlnProPheLysGlyAlaArgIleThrGlySerLeuHisMetThrIleGlnThr 67  
 QY 199 GCGTCTCTCATGACACATCAAGCTCTTGGTGTGATGTCAGATGGCTTCCTGCAAC 258  
 Db 68 AlaValLeuIleGluThrLeuThrAlaLeuGlyAlaGluValArgTyrPysSerCysAsn 87  
 QY 259 ATCTTCTTACACAGATACAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 318  
 Db 88 IlePheSerThrGlnAspHisAlaAlaAlaIleAlaArgAspSerAlaAla----- 105  
 QY 319 GAGAACCCAGCCGCTATCCAGCTCTTCGCTGGAAGGCGGAAACACTCCAGAACTACTGG 378  
 Db 106 -----ValPheAlaTrpLysGlyGluThrLeuGlnGluTyrTrp 118  
 QY 379 GAGAACACATCCGCTCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 438  
 Db 119 TrpCysThrGluArgAlaLeuAspTrpGlyProGlyGlyGlyProAspLeuIleValAsp 138  
 QY 439 GATGGTGTGTGATGATACATCTCTCATCTCCAGAGCTTCGAA-----TTCGAA 486  
 Db 139 AspGlyGlyAspAlaThrLeuLeuIleHisGluGlyValLysAlaGluGluIlePheGlu 158  
 QY 487 ACAGCCGGTGTCTTCAGAGCAACAGAGCTGACACTCGAATCAACGCTGCTGCTTCTT 546  
 Db 159 LysThrGlyGlnValProAspProThrSerThrAspAsnProGluPheGlnIleValLeu 178  
 QY 547 GCTACACTCAACAGCTCTTCAACCAAGACCAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 606  
 Db 179 SerIleIleLysGluGlyLeuGlnValAspProLysLysLysLysLysLysLysLysLys 198  
 QY 607 ATGAACGGTGTTCGAGAGACAAACAGAGTGTCCACCGCTCTTACCAGCTTCGAGAG 666  
 Db 199 LeuValGlyValSerGluThrThrThrGlyValLysArgLeuTyrGlnMetGlnGln 218  
 QY 667 GAGGCAAACTCTCTCCAGCATCAACGCTCAACGCTGCTGCTGCTGCTGCTGCTGCTGCT 726  
 Db 219 AsnGlyThrLeuPheProAlaIleAsnValAsnAspSerValThrLysSerLysPhe 238  
 QY 727 GATAACATCTACGGTCTCGCGCTCTCTCTATCGATGATCAACCGCTGCTTCCGATGTC 786  
 Db 239 AspAsnLeuTyrGlyCysArgHisSerLeuProAspGlyLeuMetArgAlaThrAspVal 258  
 QY 787 ATGATCGCGGCAAGACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 846  
 Db 259 MetIleAlaGlyLysValAlaValIleCysGlyTyrGlyAspValGlyLysGlyCysAla 278  
 QY 847 CAATCTCTCGCGGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 906  
 Db 279 AlaAlaMetLysThrAlaGlyAlaArgValIleValThrGluIleAspProIleCysAla 298  
 QY 907 CTCGAGGCTGCATGGAAGGTACAGGCTCCGCGCTCCGCGCTGAGGAGTCTGCAAGATGTC 966  
 Db 299 LeuGlnAlaLeuMetGluGlyLeuGlnValLeuThrLeuGluAspValValSerGluAla 318  
 QY 967 GATATCTCTGTTACATGACAGGAACTGCGATATCATCTCTGTCATGATGTCGCCAG 1026  
 Db 319 AspIlePheValThrThrThrGlyAsnLysAspIleIleMetValAspHisMetArgLys 338  
 QY 1027 ATGAAGCAATGAGCTATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1086  
 Db 339 MetLysAsnAlaIleValCysAsnIleGlyHisPheAspAsnGluIleAspMetLeu 358  
 QY 1087 GGCCTCATGAATACCCAGGATCAGACATCCCAATCAAGCCAGCAATACAGCATGTGG 1146

Db 359 GlyLeuGluThrTyrProGlyValLysArgIleThrIleLysProGlnThrAspArgTrp 378  
 QY 1147 GAATTTCCAGATGCGCAGCT---ATCTCTCTTCTGCTGAGGCGCGCTTCTTAACCTT 1203  
 Db 379 ValPheProGluThrIleAlaGlyIleValLeuAlaGluGlyArgLeuMetAsnLeu 398  
 QY 1204 GGTCTGCTACAGGTCACCCATCTTCTGTTATGTCATGTCATTCATTCACAAACACAGACCTC 1263  
 Db 399 GlyCysAlaThrGlyHisProSerPheValMetSerCysSerPheThrAsnGlnValIle 418  
 QY 1264 GCTCAGCTGCACTCTACGAAAGAGA-----GGAATCTCGAGAGAGGTTTACACA 1317  
 Db 419 AlaGlnLeuGluLeuTrpAsnGluLysAlaSerGlyLysTyrGluLysValTyrVal 438  
 QY 1318 CTTCGAGCATCTCGATGAGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1377  
 Db 439 LeuProLysHisLeuAspGluLysValAlaLeuLeuHisLeuGlyLysLeuGlyAlaArg 458  
 QY 1378 CTTCACAAAGTTACACAGAGCGCTGACTCATCATCAAGTTCAGGTCAGGTCCTTAC 1437  
 Db 459 LeuThrLysLeuSerLysAspGln\*\*\*AspTyrValSerIleProIleGluGlyProTyr 478  
 QY 1438 AAGTCTCATGCTTACCGTTAT 1458  
 Db 479 LysProProHisTyrArgTyr 485  
 RESULT 4  
 ID Q8LPS8 PRELIMINARY; PRT; 485 AA.  
 AC Q8LPS8;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE AT3G23810/MW9.15.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosida II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,  
 RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,  
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.B., Sakurai T., Satou M.,  
 RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
 RA Ecker J.R.;  
 RT "Arabidopsis cDNA clones";  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 EMBL: AY094404; AAM19782.1; -;  
 DR InterPro; IPR000043; Ado\_hcyase.  
 DR Pfam; PF00670; AdoHcyase; 1.  
 DR TIGRFAMs; TIGR00936; ahcy; 1.  
 DR PROSITE; PS00738; ADHCHYASE 1; 1.  
 DR PROSITE; PS00738; ADHCHYASE 2; 1.  
 SQ SEQUENCE 485 AA; 53101 MW; 14275615D0996059 CRC64;  
 Alignment Scores:  
 Pred. No.: 1.57e-103 Length: 485  
 Score: 1512.00 Matches: 299  
 Percent Similarity: 73.72% Conservative: 60  
 Best Local Similarity: 61.40% Mismatches: 112  
 Query Match: 55.24% Indels: 16  
 DB: 10 Gaps: 4  
 US-09-759-990A-1 (1-1461) x Q8LPS8 (1-485)  
 QY 19 ACTGGTCTCCATCGAGTACAGATTGCCGACATCAACCTCCATGTTCTCGCGCGTAAG 78  
 Db 8 ThrSerSerGlyArgGluTyrLysValLysAspMetSerGlnAlaAspPheGlyArgLeu 27

QY	1147	GAATTC	CGAGATGCGCAGCT	-----ATTCCTCTCTTCTCTGAGGGCGCCTCTTAACTT	1203			
Db	379	ValPheProAspThr	AsnSerGlyIleIleValLeuAlaGluGlyArgLeuMetAsnLeu	398				
QY	1204	GGCTGG	CGCTACAGGTCACCCATCTTTTCGTTATGTCAATGTCAATTCACAAACACAGACACTC	1263				
Db	399	GlyCysAlaThrGlyHisProSerPheValMetSerCysSerPheThrAsnGlnValIle	418					
QY	1264	GCTCAGCT	CGACTCTTACGAAAGAGA-----GGAAATCTCGAGAGAAGGTTTACACA	1317				
Db	419	AlaGlnLeuGluLeuTrpAsnGluLysSerSerGlyLysTyrGluLysLysValTyrVal	438					
QY	1318	CTTCCGA	AGCATCTCGATGAGGAAGTCGCTCGCCCTCCACCTCGCATCTCTCGCATGTCCAC	1377				
Db	439	LeuProGlyHisLeuAspGluLysValAlaAlaLeuHisLeuGlyLysLeuGlyAlaArg	458					
QY	1378	CTTACAAA	AGCTTACACAGACAGCGTGCATCATCAACAGTTCCAGTTGAGGGTCTCTTAC	1437				
Db	459	LeuThrLysLeuThrLysAspGlnSerAspTyrValSerIleProValGluGlyProTyr	478					
QY	1438	AAGTCT	GATGCTTACCGTTAT	1458				
Db	479	LysProValHisTyrArgTyr	485					
RESULT 5								
Q944KS	Q944KS	PRELIMINARY;	PRT;	485 AA.				
AD	Q944KS							
DT	01-DEC-2001	(T-EMBLrel. 19, Created)						
DT	01-DEC-2001	(T-EMBLrel. 19, Last sequence update)						
DT	01-OCT-2002	(T-EMBLrel. 22, Last annotation update)						
DE	AT323810/MY9_15	(EC 3.3.1.1) (Adenosylhomocysteinase) (S-adenosyl-L-						
DE	homocysteine hydrolase) (AdoCysase).							
OS	Arabidopsis thaliana	(Mouse-ear cress).						
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;							
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;							
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.							
ON	NCBI_TaxID=3702;							
RX	[1]							
RP	SEQUENCE FROM N.A.							
RA	Chew R., Chen H., Kim C.J., Koese E., Meyers M.C., Banh J.,							
RA	Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,							
RA	Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,							
RA	Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,							
RA	Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,							
RA	Saton M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,							
RA	Yamamura F., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,							
RA	Becker J.R.;							
RT	"Arabidopsis cDNA clones";							
RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.							
CC	!- FUNCTION: ADENOSYLHOMOCYSTEINE IS A COMPETITIVE INHIBITOR OF S-							
CC	ADENOSYL-L-METHIONINE-DEPENDENT METHYL TRANSFERASE REACTIONS;							
CC	THEREFORE ADENOSYLHOMOCYSTEINASE MAY PLAY A KEY ROLE IN THE							
CC	CONTROL OF METHYLATIONS VIA REGULATION OF THE INTRACELLULAR							
CC	CONCENTRATION OF ADENOSYLHOMOCYSTEINE (BY SIMILARITY).							
CC	!- CATALYTIC ACTIVITY: S-ADENOSYL-L-HOMOCYSTEINE + H(2)O = ADENOSINE							
CC	+ L-HOMOCYSTEINE.							
CC	!- COFACTOR: NAD (BY SIMILARITY).							
CC	!- PATHWAY: ACTIVATED METHYL CYCLE.							
CC	!- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYSTEINASE FAMILY.							
DR	EMBL; AF28329; AAL16259.1; -							
DR	InterPro; IPR000043; Ado_hcyase.							
DR	Pfam; PF00670; AdoHcyase; 1.							
DR	TIGRFAMs; TIGR00936; ahcy; 1.							
DR	PROSITE; PS00738; ADOHCYASE 1; 1.							
DR	PROSITE; PS00739; ADOHCYASE 2; 1.							
KW	Hydrolase; NAD; One-carbon metabolism.							
SQ	SEQUENCE 485 AA; 53129 MW; 8AA4719B3P6BFAF6 CRC64;							
Alignment Scores:								
Pred. No.:		1.57e-103	Length:	485				
Score:		1512.00						

Alignment Scores:

Pred. No.:	1.57e-103	Length:	485
Score:	1512.00	Matches:	299
Percent Similarity:	73.73%	Conservative:	60



RP SEQUENCE FROM N.A.  
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,  
RA Palm C.J., Bowser L., Jones T., Barth J., Carninci P., Chen H.,  
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kaniya A., Kawai J.,  
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
RA Sakurai T., Satou M., Seki M., Shim W., Yamada K., Shinozaki K.,  
RA Ecker J., Theologis A., Davis R.W.,  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,  
RA Quach H.L., Tang C.C., Toriumi M., Wallender B.K., Wong C., Wu H.C.,  
RA Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J., Nguyen M.,  
RA Palm C.J., Shinn P., Southwick A., Tripp M.G., Wu T., Davis R.W.,  
RA Ecker J.R., Theologis A.,  
RT "Arabidopsis Open Reading Frame (ORF) Clones";  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ADENOSYLHOMOCYSTEINE IS A COMPETITIVE INHIBITOR OF S-  
CC ADENOSYL-L-METHIONINE-DEPENDENT METHYL TRANSFERASE REACTIONS;  
CC THEREFORE ADENOSYLHOMOCYSTEINASE MAY PLAY A KEY ROLE IN THE  
CC CONTROL OF METHYLATIONS VIA REGULATION OF THE INTRACELLULAR  
CC CONCENTRATION OF ADENOSYLHOMOCYSTEINE (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-HOMOCYSTEINE + H(2)O = ADENOSINE  
CC + L-HOMOCYSTEINE.  
CC -!- COFACTOR: NAD (BY SIMILARITY).  
CC -!- PATHWAY: ACTIVATED METHYL CYCLE.  
CC -!- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYSTEINASE FAMILY.  
DR EMBL: AP000377; BAB01858.1; -.  
DR EMBL: AY059888; AAL24370.1; -.  
DR EMBL: AY093385; AAM13384.1; -.  
DR EMBL: AY150471; AAN12996.1; -.  
DR HSP: P10760; IBSR.  
DR InterPro: IPR000043; Ado\_hcyase.  
DR Pfam: PF00670; Adohcyase; 1.  
DR TIGRFAMs: TIGR00936; ahcy; 1.  
DR PROSITE: PS00738; ADOHCYASE\_1; 1.  
DR PROSITE: PS00739; ADOHCYASE\_2; 1.  
KW Hydrolase; NAD; One-carbon metabolism.  
SQ SEQUENCE 485 AA; 53159 MW; 876079E944DC3732 CRC64;

Alignment Scores:  
Pred. No.: 1.57e-103 Length: 485  
Score: 1512.00 Matches: 299  
Percent Similarity: 73.72% Conservative: 60  
Best Local Similarity: 61.40% Mismatches: 112  
Query Match: 55.24% Indels: 16  
DB: 10 Gaps: 4

US-09-759-990A-1 (1-1461) x Q9LK36 (1-485)

QY 19 ACTGGTGTCATTCGAGTACAGAAATGCCGACATCACTCCATGTTCTCGCGCGTAAG 78  
Db 8 ThrserserGlyArggluTyrLysValLysAspMetSerGlnAlaAspPheGlyArgLeu 27  
QY 79 GAATCTACCTTGCTGAGAGGAAATGCCAGGTCTTATGAGTCTTCGTGAGGCGTTATTCC 138  
Db 28 GluileGluLeuAlaGluValGluMetProGlyLeuValSerCysValThrGluPheGly 47  
QY 139 GCTTCTAAGCCATTGAGGGGTGTCAGAAATCTGTGTTCCCTCCACATGACATGCCAGACA 198  
Db 48 ProSerGlnProLeuLysGlyAlaArgileThrGlySerLeuHisMetThrileGlnThr 67  
QY 199 GCGTCTCTCATCGACACACTCACAGTCTTGTGCTGATGTCAGATGGGCTTCCTGCAAC 258  
Db 68 AlaValleulleGluThrLeuThrAlaLeuGlyAlaGluValAlaArgirpCysSerCysasn 87  
QY 259 ATCTTCTCFACCAAGATACAGCGCTGCTGCTATCTGTTGTCGGCCCAACAGCGCACCA 318  
Db 88 IlePheSerThrGlnAspHisAlaAlaAlaAlaileAlaAlaArgAspSerAlaAla----- 105  
QY 319 GAGAGCGACCGCGTATCCCGAGTCTTCGCTCGAGGGCGGAAACACTCCAGAAATCTGG 378  
Db 106 -----ValPheAlaTrpLysGlyGluThrLeuGlnGluTyrTrp 118

QY 379 GAGAACACATACCGCGCTCTCCATGCGGCACATGTCGTCAGAGCCACACAGGTTGTGGAT 438  
Db 119 TrpCysThrGluArgAlaLeuAspTrpGlyProGlyGlyGlyProAspLeuileValAsp 138  
QY 439 GATGGTGTGATGCTACACTCTCTCTATCTCCAAAGGGCTTCGAA-----TTCGAA 486  
Db 139 AspGlyGlyAspAlaThrLeuileHisGluGluValLysAlaGluGluilePheAla 158  
QY 487 ACACGCGGTCTCTCCAGACCAACAGAGCTGCAACACTCCGAATACCGTGCGTCTCTT 546  
Db 159 LysasnGlyThrPheProaspProThrSerThrAspAsnProGluPheGlnIleValLeu 178  
QY 547 GCTACACTCAAGCAGTCTTCAACACACAGAACACACTGCCACACAGTGTCTGCGGCG 606  
Db 179 SerileileLysAspGlyLeuGlnValAspProLysLysTyrrHisLysMetLysLysGluArg 198  
QY 607 ATGAACGCTGTTCCGAGACGACACACAGGTTCCACCGCTCTACACGCTCGAGAG 666  
Db 199 LeuValGlyValSerGluGluThrThrGlyValLysArgLeuTyrrGlnMetGlnGlu 218  
QY 667 GAGGCGAAACTCTCTTCCAGCCATCAACGTCACAGCAGCTGTTCACAAAGTCCAAATTC 726  
Db 219 ThrGlyAlaLeuLeuPheProAlaileasnValasnAspSerValThrLysSerLysPhe 238  
QY 727 GATACACTACGCTGCGGCACCTCTCTATGATGATGATCAACCGTCTCGATGTC 786  
Db 239 AspAsnLeuTyrrGlyCysArgHisSerLeuProAspGlyLeuMetArgAlaThrAspVal 258  
QY 787 ATGATCGCGCGGCAAGACAGTCTCTGTCATGGTTACGGGATGTCGCGCAAGGCTGGCT 846  
Db 259 MetilealagLysValAlaValleileCysGlyTyrrGlyAspValGlyLysGlyCysAla 278  
QY 847 CAATCCCTCGTGGCCAGCGCTCGGTATCATCACAGAACTCGACCCCAATCTGGCT 906  
Db 279 AlaalaMetLysThrAlaGlyAlaArgValleileValThrGluileAspProileCysAla 298  
QY 907 CTCGAGCTGCCATGAGGCTACCGAGTCCGCGGCTCGAGGAAGTCTCAAGATGTC 966  
Db 299 LeuGlnAlaLeuMetGluGlyLeuGlnValleuThrLeuGluAspValValSerGluAla 318  
QY 967 GATATCTTCTGTATACACAGGAACTGGCATATCATCTCTGTTCATGATGCGCCAG 1026  
Db 319 AspIlePheCysThrThrThrGlyasnLysAspIleileMetValAspHisMetArgLys 338  
QY 1027 ATGAAGGATAGGCTATTCTCGTAACTACGCGCCTTCGATAACGAAATTCATACAGAT 1086  
Db 339 MetLysAsnAsnAlaileValCysAsnileGlyHisPheAspAsnGluileAspMetLeu 358  
QY 1087 GCGCTCATGAAATACCCAGGCATCAAGCACATCCCAATCAAGCCAGATACGATGTGG 1146  
Db 359 GlyLeuGluThrTyrrProGlyValLysArgileThrileLysProGlnThrAspAspGtrp 378  
QY 1147 GAATTCACGATGCCACGCT---ATCTCTCTTCTGCTGAGGGCGCGCTTCTTAACCTT 1203  
Db 379 ValPheProaspThrAsnSerGlyileileValleuAlaGluGlyArgLeuMetAsnLeu 398  
QY 1204 GCGTCCGCTACAGTCCACCATCTTCGTTATGTCATGTCATGTCATCACAACACGACACTC 1263  
Db 399 GlyCysAlaThrGlyHisProSerPheValMetSerCysSerPheThrAsnGlnValile 418  
QY 1264 GTCATGCTCCACTCTACGAAAGAGA-----GGAAATCTCGAGAGAGGTTTACACA 1317  
Db 419 AlaGlnleuGluLeuThrAsnGluLysSerSerGlyLysTyrrGluLysValTyrrVal 438  
QY 1318 CTTCCGAGATCTTCGATGAGAGTCTGCTGCTCCACTCCGATCTCTCGATGTCAC 1377  
Db 439 LeuProLysHisLeuAspGluLysValAlaAlaLeuHisLeuGlyLysGlyAlaArg 458  
QY 1378 CTTACAAAGCTTACACAGAGCGCTGACTATCATCAACGTTCCAGTTGAGGGTCTTAC 1437  
Db 459 LeuThrLysLeuThrLysAspGlnSerAspTyrrValSerileProValGluGluTyrrTrp 478

QY	274	GATACAGCCGCTGCTGTATCGTTGTTCGCCCAACAGGCACACAGAGAAGCACC	GGT	333
D	:		:	
Db	92	AspHisAlaAlaAlaAlaAlaA--ThrGlyThr-----		103
QY	334	ATCCCAGTCTTCGCTCGAGAGGGGAAACACTCCAGAACTACTGGAGAACACATACGC		393
D	:		:	
Db	104	--ProValPheAlaIatrlPySgLyGIuThrLeuGluiTrpaspCysThrLeuAsp		122
QY	394	GCTCTCACATGG-----CCAGATGGT-----CAAGGCCACAGACAGAGTTCCTCATGAT		441
D	:		:	
Db	123	AlaLeuThr-PheThr-LeuProAspGlyThrLeuThrGlyProGluLeuValValAspAsp		142
QY	442	GGTGGTGATGCTACACTCTCTCATCTCCAGGGCTTCGAATTCGAAACACCGCGTCT---		498
D	:		:	
Db	143	GlyGlyAspValThrLeuLeuIleHisLysGlyTyrgLuLeuGluAsnGlySerThrTrp		162
QY	499	GTTCACAGACCAAAGAAGCTGACAACCTCGAATAACCGCTGCTTCTCTACACTCAAG		558
D	:		:	
Db	163	ValAspGluPro-----AlaSerSerHisGluGluGlyValIleLysAlaLeuLeuLys		180
QY	559	CAGGTCTTCAACCAAGACAGAACCCTGGCACACAGTGTCTCCGGCATGAAGCGTGT		618
D	:		:	
Db	181	ArgValAlaValGluAtPrpGlyTyrrTpGlyArgValVallLysAspTrpLysGlyVal		200
QY	619	TCCGAAGACACAAACAAGGTGTCACCGCTTCACAGCTCCGAGAGGAGGAGGCAATC		678
D	:		:	
Db	201	SerGluGluThrThrThrGlyValHisArgLeutyrglnIleAlaGluAlaGlyLysLeu		220
QY	679	CTCTTCCCAGCCATCAACGCTCAACGACGCTGTTCACAAAGTCCAAAGTTCGATAACATCTAC		738
D	:		:	
Db	221	LeuIleProAlaIleAsnValAsnAspSerValThrLysSerLysPheAspAsnLeuTyrr		240
QY	739	GGCTCGCGCCACTCCCTNTTCGATGTATCAACGCTGCTTCCGATGTCTATGATCGCGCGC		798
D	:		:	
Db	241	GlyCyArgGluSerLeuAlaAspGlyLeuLysArgAlaMetAspValMetLeuAlaGly		260
QY	799	AAGACAGCTCTCTGTCATGGGTATAGCGCATGTCGCGAAGGGCTGCGCTCAATCCCTCCGT		858
D	:		:	
Db	261	LysValAlaValValCysGlyTyrgLyAspValGlyLysGlySerAlaAlaSerLeuArg		280
QY	859	GGCCAAGGCGCTCGGTTATCATCAGAACTCGACCAATCTGCGCTCTCCAGGCTGCC		918
D	:		:	
Db	281	AlatyrgLyAlaArgValIleValThrGluIleAspIleCysAlaLeuGluAlaSer		300
QY	919	ATGGAAGGCTACCAAGTCCGCGCATCGAGGAAGTCTCAAGGATGTCGATATCTTGTT		978
D	:		:	
Db	301	MetGluGlyPheGluValAsnThrIleGluSerThrLeuGlyArgAlaAspIleTyrrVal		320
QY	979	ACHTCACAGGAACCTCGATATCATCTCTGTGTACATGANGCCACAGATGAAGATAAG		1038
D	:		:	
Db	321	Thr-ThrThrGlyAsnLysAspIleThrValGluHisLeuGluAlaMetLysAspGln		340
QY	1039	GCATTATGTCGGTAATCATCGCCATCTCGATACCAAAATTCATACAGATGGCTCATGAA		1098
D	:		:	
Db	341	AlaIleValCysAsnIleGlyHisAspAsnGluIleGlnValAspAlaLeuLysAla		360
QY	1099	TACCCAGGCATCAACGACATCCCCAATCAAGCCAGAAATACGACATGTGGGAATCCCGAT		1158
D	:		:	
Db	361	LeuLysAspValGlnLysIleAsnIleLysProGlnValAspLysTyrrValPheProAsn		380
QY	1159	GGCCACGCTATCTCTCTTCTGTCTGAGGGCGCGCTCTTAACTTGTGCTGGCTACAGGT		1218
D	:		:	
Db	381	GlyAsnAlaIlePheLeuLeuAlaAspGlyArgLeuValAsnLeuGlyCysAlaThrGly		400
QY	1219	CACCCATCTTTCGTATGTCAA'TCATTCACAAACACAGACATCGCTCAGCTCGACCTC		1278
D	:		:	
Db	401	HisProSerPheValMetSerAsnSerPheAlaAsnGlnThrLeuAlaGlnIleAspLeu		420
QY	1279	TACGAAAAGAGAGGAANCTTCGAGAGAAGGTTTACACATTCGGAAGCATCTCGATGAA		1338
D	:		:	
Db	421	TrpGluLysArgAspThrTyrrGluLysLysValTyrrileLeuProLysHisLeuAspGlu		440

QY 1339 GAAGTCGCTCGCTCCACCTCGCATCTCCGATGTCCACCTTACAAAGCTTACAGAGAAG 1398  
DB 441 GluValAlaArgLeuHisLeuGluLysIleGlyValLysLeuThrLeuThrLysAsp 460  
QY 1399 CAGCGTCTACTACATCAAGCTTCAGTGGGTCTCTTACAGTCTGATCTTACCGTTAT 1458  
DB 461 GlnAlaAspTyrLeuGlyValAspValAlaGlyProTyrLysProAspHisTyrArgTyr 480  
RESULT 8  
Q94929 PRELIMINARY; PRT; 485 AA.  
AC Q94929;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Putative S-adenosyl L-homocysteine hydrolase (EC 3.3.1.1)  
DE (Adenosylhomocysteinase) (S-adenosyl-L-homocysteine hydrolase)  
DE (AdoHcyase).  
GN MYH9.16.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
SEQUENCE FROM N.A.  
RA Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Dale J.M.,  
RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,  
RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B.,  
RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,  
RA Sakurai T., Satou M., Seki M., Shimizu P., Southwick A., Tracy S.B.,  
RA Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;  
RT "Full Length cDNA of gene MYH9.16 (GI:9293955).";  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ADENOSYLHOMOCYSTEINE IS A COMPETITIVE INHIBITOR OF S-  
CC ADENOSYL-L-METHIONINE-DEPENDENT METHYL TRANSFERASE REACTIONS;  
CC THEREFORE ADENOSYLHOMOCYSTEINASE MAY PLAY A KEY ROLE IN THE  
CC CONTROL OF METHYLATIONS VIA REGULATION OF THE INTRACELLULAR  
CC CONCENTRATION OF ADENOSYLHOMOCYSTEINE (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-HOMOCYSTEINE + H(2)O = ADENOSINE  
CC + L-HOMOCYSTEINE.  
CC -!- COFACTOR: NAD (BY SIMILARITY).  
CC -!- PATHWAY: ACTIVATED METHYL CYCLE.  
CC -!- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYSTEINASE FAMILY.  
DR EMBL; AY050783; AAK92718.1; -.  
DR InterPro; IPR000043; Ado\_hcyase.  
DR Pfam; PF00670; AdoHcyase; 1.  
DR TIGRFAMs; TIGR00936; ahcy; 1.  
DR PROSITE; PS00738; ADOHCYASE\_1; 1.  
DR PROSITE; PS00739; ADOHCYASE\_2; 1.  
KW Hydrolase; NAD; One-carbon metabolism.  
SQ SEQUENCE 485 AA; 53178 MW; 876079F4782FE532 CRC64;  
Alignment Scores:  
Pred. No.: 6.17e-103 Length: 485  
Score: 1504.00 Matches: 298  
Percent Similarity: 73.51% Conservative: 60  
Best local Similarity: 61.19% Mismatches: 113  
Query Match: 54.95% Indels: 16  
DB: Gaps: 4  
US-09-759-990A-1 (1-1461) x Q94929 (1-485)  
QY 19 ACTGGTGCTCCATTGAGTACAGAAATGCGACATCAACCTCCATGTTCTCGCGCGTAAG 78  
DB 8 ThrSerSerGlyArgGluTyrLysValLysAspMetSerGlnAlaAspPheGlyArgLeu 27  
QY 79 GAACTTACCTTGCTGAGAGAAATCCAGAGTCTTATGTTCTTCGTGAGCGTTATTCC 138  
DB 28 GluIleGluLeuAlaGluValGluMetProGlyLeuValSerCysValThrGluPheGly 47

QY 1339 GCTTCTAAGCCATTGAAGGGTGTCCAGAAATCTCTGGTTCCTCCATCCAGTCCAGACA 198  
DB 48 ProSerGlnProLeuLysGlyAlaArgIleThrGlySerLeuHisMetThrIleGlnThr 67  
QY 139 GCGTCTCTATCGAGACTCAGAGCTCTTGGTGTCTGATGTCAGATGGGCTTCTCTCAAC 258  
DB 68 AlaValLeuIleGluThrLeuThrAlaLeuGlyAlaGluValArgTyrCysSerCysAsn 87  
QY 259 ATCTTCTCTACCAAGATACAGCGCTGCTGCTATCTGTTGCGGCCCAACAGGACACCA 318  
DB 88 IlePheSerThrGlnAspHisAlaAlaAlaAlaAlaAlaArgAspSerAlaAla----- 105  
QY 319 GAGAAGCCAGCGGTATCCCGTCTTCCCTGGAAGGGCGAAACACTCCCGAATATCTGG 378  
DB 106 -----ValPheAlaTrpLysGlyIleThrLeuGlnGluTyrTrp 118  
QY 379 GAGAACACATACCGCGCTCTCACATGCCAGATGGTCAAGGCCACAGCAGAGTGTTCGAT 438  
DB 119 TrpCysThrGluArgAlaLeuAspTrpGlyProGlyGlyGlyProAspLeuIleValAsp 138  
QY 439 GATGTTGTTGATGCTACTCTCTCATCTCCAGGGGTTTCGAA-----TTCGAA 486  
DB 139 AspGlyGlyAspAlaThrLeuLeuIleHisGluGlyValLysAlaGluGluIlePheAla 158  
QY 487 ACAGCCGCTGCTTCCAGAGCCAGAGAGCTGACAACTCGAATACCGTCTCGTCTT 546  
DB 159 LysAsnGlyThrPheProAspProThrSerThrAspAsnProGluPheGlnIleValLeu 178  
QY 547 GCTACACTCAAGCAGGTCTTCAACCAAGAACCAACTGCGACACAGATTGCTGCGCGC 606  
DB 179 SerIleIleLysAspGlyLeuGlnValAspProLysTyrHisLysMetLysGluArg 198  
QY 607 ATGAACGGTGTTCGAGAGACAAACAGGTGTCCACCGCTCTACAGCTCGAGAGAG 666  
DB 189 LeuValGlyValSerGluGluThrThrGlyValLysArgLeuTyrGlnMetGlnGlu 218  
QY 667 GAGGCAAACTCTCTTCCAGCATCAAGCTCAAGCAAGCTGTTTACAAAGTCCAAAGTTC 726  
DB 219 ThrGlyAlaLeuLeuPheProAlaAlaAsnValAspSerValThrLysSerLysPhe 238  
QY 727 GATAACATACGGTTCGCGCACTTCCCTTATCGATGATCAACCGTCTCCGATGTC 786  
DB 239 AspAsnLeuTyrGlyCysArgHisSerLeuProAspGlyLeuMetArgAlaThrAspVal 258  
QY 787 ATGATCGCGCGGAGAGACAGCTCTCGTCATGGTTCAGCGATGTCGCGAGGGTTCGCT 846  
DB 259 MetIleAlaGlyLysValAlaValIleCysGlyTyrGlyAspValGlyLysGlyCysAla 278  
QY 847 CAATCTCTCGTCCGCAAGCGCTCGCGTATCATCACAGAACTCGACCAACTCGCGCT 906  
DB 279 AlaAlaMetLysThrAlaGlyAlaArgValIleValThrGluIleAspProIleCysAla 298  
QY 907 CTCAGAGCTCCATGGAAGGTACAGGTTCGCGCATCGAGAAAGTCTCAGAGGATGTC 966  
DB 299 LeuGlnAlaLeuMetGluGlyLeuGlnValLeuThrLeuGluAspValValSerGluAla 318  
QY 967 GATATCTCTGTATCATGACAGGAACTCGCGTATCATCTCTGTGTGATCATGTCGCGCAG 1026  
DB 319 AspilePheCysThrThrThrGlyAsnLysAspIleIleMetValAspHisMetArgLys 338  
QY 1027 ATGAAGGATAGCGTATTCGTGTAATCGGCACTTCGATAAGAAATGATACAGAT 1086  
DB 339 MetLysAsnAsnAlaIleValCysAsnIleGlyHisPheAspAsnGluIleAspMetLeu 358  
QY 1087 GCGTCTATGAATACCCAGGATCAAGCATCCCAATCAAGCCAGAAATACAGATGCG 1146  
DB 359 GlyLeuGluThrTyrProGlyValLysArgIleThrIleLysProGlnThrAspArgTrp 378  
QY 1147 GAAATTCACGATGCCCACT 1203  
DB 379 ValPheProAspThrAsnSerGlyIleIleValLeuAlaGluGlyArgLeuMetAsnLeu 398  
QY 1204 GGCTCGCGTACAGGTCAACCCATCTTCTGTTATGTGTAATGTCTTCAACACAGCACTC 1263



DQ		301	METGILGIFYRHEGIIVAVASNNIILLEGUWSEINLEGGSYAAGSGYSPFARRELYVAL	326
QY		979	ACATGCACAGGAACAATCGCATPATCATCTCTGTGTGACATGATGCCCCAGATGAAGAAGTAAG	1038
DB		321	Thr-Thr-GlyAspLysAspIleIleThrValGluHisLeuGlnAlaMetIysAspGln	340
			::::	
			::::	
QY		1039	GCTATTGTTCGGTAACTACATCGGCCACTTCGATACGGAATTGATACAGATGGCCTCATGAAA	1098
DB		341	AlaIleValCysAsnIleGlyHISpeAspAsnGluIleGlnValaspAlaLeuasnla	360
QY		1099	TACCACGGCATCAAGCACATCCCACATCAAGCCACGAATACGACATGTGGGAANTCCACAGAT	1158

Alignment Scores:	
Pred. No.:	1.22e-102
Score:	1500.00
Length:	480
Percent Similarity:	75.00%
Matches:	306
Best Local Similarity:	63.75%
Conservative:	54
Query Match:	104
Mismatches:	16
Indels:	16
DB:	Gaps: 6
US-09-759-990A-1 (1-1461) x Q8PCH5 (1-480)	



```

Db      361 LeuLysGlyValGluLysIleAsnIleLysProGlnValAspLysTyrValPheGlyAsn 380
QY      1159 GGCACGCTATCTCTTCTGCTAGGCGCGCTTCTTAACCTTGGCTGCGCTACAGGT 1218
Db      381 GlyAsnAlaIlePheLeuLeuAlaAspGlyArgLeuValAsnLeuGlyCysAlaThrGly 400
QY      1219 CACCATCTTTCGTTATGTCAATTCACAAACAGACACTCCGCTCAGCTCGACCTC 1278
Db      401 HisProSerPheValMetSerAsnSerPheAlaAsnGlnThrLeuAlaGlnIleAspLeu 420
QY      1279 TACGAAAAGAGAGAAATCTCGAAGAAGTTTACACTTCGGAAGCATCTCATGAA 1338
Db      421 TrpGluLysArgAspSerTyrGluLysLysValTyrIleLeuProLysHisLeuAspGlu 440
QY      1339 GAAGTCGCTCGCTCCACTCGGATCTCTCGATGCCACTTACAAAGCTTACACAGAAG 1398
Db      441 GluValAlaArgLeuHisLeuGluLysIleGlyValLysLeuThrThrLeuThrLysAsp 460
QY      1399 CAGGCTGACTACATCAAGCTTCCAGTTGAGGCTCTTACAAAGTCTGATCTTACCGTTAT 1458
Db      461 GlnAlaAspTyrLeuGlyValAspValAlaGlyProTyrLysProAspHisTyrArgTyr 480

RESULT 10
Q8FRJ4 PRELIMINARY; PRT; 478 AA.
AC      Q8FRJ4;
DT      01-MAR-2003 (TrEMBLrel. 23, Created)
DT      01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE      Putative adenosylhomocysteinease.
GN      CE0767.
OS      Corynebacterium efficiens.
OC      Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC      Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX      NCBI_TaxID=152794;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA      Kawarabayashi Y., Yamazaki J., Hano Y., Kikuchi H., Nakamura Y.,
RA      Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA      Usuda Y., Sugimoto S.;
RT      "The entire genomic sequence of Corynebacterium efficiens YS-314."
RL      Submitted (MAY-2002) to the EMBL/GenBank/DBSJ databases.
DR      EMBL; AP005216; BAC17577.1; -.
KW      Complete proteome.
SQ      SEQUENCE 478 AA; 52776 MW; A75845A1A3F3A978 CRC64;

Alignment Scores:
Pred. No.: 2,21e-101 Length: 478
Score: 1483.00 Matches: 283
Percent Similarity: 76.00% Conservative: 78
Best Local Similarity: 59.58% Mismatches: 112
Query Match: 54.18% Indels: 2
DB: 16 Gaps: 2

US-09-759-990A-1 (1-1461) x Q8FRJ4 (1-478)
QY      34 GAGTACAGAAATCGCGACATCAACCTCCATGTTCTCGCGGTAAAGNACTTACCTTCT 93
Db      6 AspPheLysValAlaAspLeuSerLeuAlaGluAlaGlyArgHisGlnIleArgLeuAla 25
QY      94 GAGAAGAAATCCAGGCTTATGTTCTTCGTAGCGGTATTCGCGCTTCAAGCCATTG 153
Db      26 GluTyrGluMetProGlyLeuMetGlnLeuArgArgGluTyrAlaGluGluGlnProLeu 45
QY      154 AAGGTTGTGAGAAATCTGTTCTCTCCATGACAGTCGACAGCGCTCTCATCGAG 213
Db      46 LysGlyAlaArgIleAlaGlySerIleHisMetThrValGlnThrAlaValLeuIleGlu 65
QY      214 ACATCTCAGCTTGTGTTGCTGATGTCAGATGGCTTCTTCCACATCTTCTCTACAA 273
Db      66 ThrLeuThrAlaLeuGlyAlaGluValArgTrpAlaSerCysAsnIlePheSerThrGln 85

```

```

QY      274 GATACAGCGCTGCTGCTATGTTGCTGGCCCAACAGGACACAGGAGAACCCAGCGGT 333
Db      86 AspGluAlaAlaAlaIleValValGly---AspGlyThrProGluAspProGlnGly 104
QY      334 ATCCAGCTCTTGGCTTGAAGGCGAAACACTCCAGATACTCCGAGAAATCTGGGAAACACATACCG 393
Db      105 ValProValPheAlaThrLysGlyGluThrLeuAspGluTyrTyrTrpCysIleAsnGln 124
QY      394 GCTCTCATGCGCAGATGCTCAAGGCCACAGCAGGTGTGCGATGATGCTGCTGATGCT 453
Db      125 IlePheSerTrp---GluGlyGluLeuProAsnMetIleLeuAspAspGlyGlyAspAla 143
QY      454 ACATCTCTCATCTCCAAAGCGCTTCGAATTCGAAACAGCGGTGCTTCCAGAGCCACA 513
Db      144 ThrMetAlaValIleArgGlyValGluTyrGluLysAlaGlyValValProGlnProGlu 163
QY      514 GAAGCTGACACCTCGAATCCGCTGCTTCTGCTTACACTCAAGCAGGTCTTCAACCAA 573
Db      164 AlaAsnAspSerAspGluTyrIleAlaPheLeuGlyMetLeuArgGluValLeuAlaGlu 183
QY      574 GACAAGAACACCTGGCACACAGTTGCTCGCGCATGAACGGTGTTCGGAAGAGACAACA 633
Db      184 GluProAspLysTrpThrArgLeuAlaAspSerIleLysGlyValThrGluGluThr 203
QY      634 ACAGTCTCCAGCGCTCTTACAGCTCGAAGAGGGCAAACTCTCTTCCAGCCATC 693
Db      204 ThrGlyValHisArgLeuTyrHisPheAlaGluGlyValLeuProPheProAlaMet 223
QY      694 AACCTCAACGAGCTGTACAAAGTCCAAAGTTCGATTAACATCTACGCTCCGCCACTCC 753
Db      224 AsnValAsnAspAlaValThrLysSerLysPheAsnLysTyrGlyThrArgHisSer 243
QY      754 CTTATCGAGTATCAACCCGCTTCCGATGTCATGCGCGGCAAGACAGCTCTCGTC 813
Db      244 LeuIleAspGlyIleAsnArgAlaThrAspMetLeuMetGlyGlyLysAsnValLeuVal 263
QY      814 ATGGGTTACGGGATGTCGGCAAGGCTGGCTCAATCCCTCCGTCGGCAAGGCTCCG 873
Db      264 CysGlyTyrGlyAspValGlyLysGlyCysAlaGluAlaPheAspGlyGlnGlyAlaArg 283
QY      874 GTTATCATCAGAACTCGACCCCAATCTGGCTCTCTCCAGCTGCGCAGAGGTACAG 933
Db      284 ValArgValThrGluAlaAspProIleAsnAlaLeuGlnAlaLeuMetAspGlyTyrSer 303
QY      934 GTCGCGCGATCGAGGAAGTGTCAAGATGTCGATATCTTCGTTATCATGCACAGAAAC 993
Db      304 ValValThrValAspGluAlaIleAlaAspAlaAspIleValIleThrAlaThrGlyAsn 323
QY      994 TCGATATCATCTCTGTCGATGATGCGCCAGATGAAGGATTAAGGCTATTGTCGTAAC 1053
Db      324 LysAspIleIleSerTyrGluGlnMetLeuLysMetLysAspHisAlaLeuLeuGlyAsn 343
QY      1054 ATCGGCCACTTCGATAAGAAATTGATACAGATGGCGCTCATGAATATCCAGGCATCAAG 1113
Db      344 IleGlyHisPheAspAsnGluIleAspMetHisSerLeuLeuHisArgAspAspValIle 363
QY      1114 CACATCCCAATCAAGCAGATACAGATGTCGGAATTCACAGATGCCAGCTATCCCTC 1173
Db      364 ArgThrThrIleLysProGlnValAspGluPheThrPheProAsnGlyLysSerIleIle 383
QY      1174 CTTTCTGCTGAGGCGCGCTCTTAACTTGGCTGCGCTTACAGGTCCACCATCTTTCGTT 1233
Db      384 ValLeuSerGluGlyArgLeuLeuLeuAsnLeuGlyAsnAlaThrGlyHisProSerPheVal 403
QY      1234 ATGTCAATGTCTTACAAACGACACTCGCTCAGCTCGACTCGACTTACGAAAGAGAGA 1293
Db      404 MetSerThrSerPheAlaAspGlnThrIleAlaGlnIleGluLeuPheGlnAsnGluGly 423
QY      1294 AATCTCGAGAAGAAGGTTTACACACTTCCGAAGCATCTCGATGAAGAGTCCGTCGCTC 1353
Db      424 GlnTyrGluAsnGlnValTyrArgLeuProLysIleLeuAspGluLysValAlaArgIle 443

```

QY	1354	CACCTCGAGATCTTCGATGTGCCACTTCAAGAAGCTTTACAAAGCAGGAGCTGACTACATC	1413
Ds	444	HISvalAlaLeuGLyGLyLysLeuThrGluLeuThryLysGlnGluAlaGluTryle	463
QY	1414	AAGCTTCCAGTTGAGGGGCCTTACAAGCTTGATGCTTACCCTTAT	1458
Ds	464	GlyValAspValAlaGlyProPheLysProGluHisTyraIgyr	478
RESULT 11			
Q42939		PRELIMINARY; PRT; 450 AA.	
ID	Q42939		
QC	Q42939;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, last sequence update)		
DT	01-OCT-2002 (TrEMBLrel. 22, last annotation update)		
DE	Cytokinin binding protein CBP57 (EC 3.3.1.1) (Adenosylhomocysteinase)		
DE	(S-adenosyl-L-homocysteine hydrolase) (AdoHcyase).		
OS	Nicotiana glauca (Wood tobacco).		
OS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
OC	Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.		
OX	NCBI_taxid=4096;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Mitsui S., Wakaugi T., Sugitara M.;		
RT	"A cDNA encoding the 57 kDa subunit of a cytokinin-binding protein complex from Tobacco : the subunit has high homology to S-adenosyl-L-		
RT	homocysteine hydrolase.";		
RL	Plant Cell Physiol. 34:1089-1096(1993).		
CC	-!- FUNCTION: ADENOSYLHOMOCYTEINE IS A COMPETITIVE INHIBITOR OF S-		
CC	ADENOSYL-L-METHIONINE-DEPENDENT METHYL TRANSFERASE REACTIONS;		
CC	THEREFORE ADENOSYLHOMOCYTEINASE MAY PLAY A KEY ROLE IN THE		
CC	CONTROL OF METHYLATIONS VIA REGULATION OF THE INTRACELLULAR		
CC	CONCENTRATION OF ADENOSYLHOMOCYTEINE (BY SIMILARITY).		
CC	-!- CATALYTIC ACTIVITY: S-ADENOSYL-L-HOMOCYTEINE + H(2)O = ADENOSINE		
CC	+ L-HOMOCYTEINE.		
CC	-!- COFACTOR: NAD (BY SIMILARITY).		
CC	-!- PATHWAY: ACTIVATED METHYL CYCLE.		
CC	-!- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYTEINASE FAMILY.		
DR	EMB1; D16139; BAA03710.1; --.		
DR	HSP; P10760; IJ3R.		
DR	InterPro; IPR000043; Ado_hcyase.		
DR	Ffam; PF00670; Adohcyase; 1.		
DR	TIGRFAMs; TIGR00936; ahcy; 1.		
DR	PROSITE; PS00738; ADOHCYASE 1; 1.		
DR	PROSITE; PS00739; ADOHCYASE 2; 1.		
KW	Hydrolase; NAD; One-carbon metabolism.		
SQ	SEQUENCE 450 AA; 49258 MW; B4B0108AF537DAE CRC64;		
Alignment Scores:			
Pred. No.:	8,47e-101	Length:	450
Score:	1475.00	Matches:	292
Percent Similarity:	74.95%	Conservative:	52
Best Local Similarity:	63.62%	Mismatches:	99
Query Match:	53.89%	Indels:	16
DB:	10	Gaps:	4
US-09-759-990A-1 (1-1461) x Q42939 (1-450)			
QY	103	ATGCCAGGCTTATTGGTCTTCGTCGAGCGGTATTCCGGCTTTTAAGCCATTGAAGGGTGTC	162
Ds	1	MetProGlyLeuMetAlaCysArgThrGluPheGlyProSerGlnProPheLysGlyAla	20
QY	163	AGATCTCTGTTCCCTCCCATGACATGACAGTCCAGACAGCGCGTCCTCATCGACAGACTCAC	222
Ds	21	LysileThrGlySerLeuHisMetThrIleGlnThrAlaValLeuIleGluThrLeuThr	40
QY	223	GCTCTCGTCTGCTGATGATGGGCTTCCTGCACCATCTTCTCTACACAGATACAGCC	282
Ds	41	AlaLeuGlyAlaGluValArgProCysSerCysAsnIlePheSerThrGlnAspHisAla	60
QY	283	GCTGCTGCTATCTTGTTCGGGCCCAACAGGCACACCAAGAACGCCAGCCGCGGTATCCCACTC	342

61	AlaAlaAlaIleAlaArgSerArgAla-----Val	71
343	TTCCCTCGAAGGGCGAAACACTCCAGAAATACTGGGAAACACATACCGCTCTCAC	402
72	PheAlaTrpLysGlyGluThrLeuGlnGluTyrTrpCysThrGluArgAlaLeuAsp	91
403	TGCCAGATGTCAAAGGCCACAGCAGGTTCGATGATCGTGATCGTGATCTACCTCTC	462
92	TrpGlyProGlyGlyGlyProAspLeuIleValAspAspGlyGlyAspAlaThrLeuLeu	111
463	ATCTCCAAAGGGCTTC-----GAATTCGAAACAGCCGGTCTGTTCACAGCCA	510
112	IleIleGluGlyValLysAlaGluGluTyrAlaLysSerGlyLysLeuProAspPro	131
511	ACAGAAGTCACAACCTCGAATACCGCTCGTTCCTGCTACATCACTCAAGCAGGTCTTCAAC	570
132	SerSerThrAspAsnAlaGluPheGlnLeuValLeuThrIleIleArgAspGlyLeuLys	151
571	CAAGACAAGAACCACTGGCACACAGTTGCTCGCGCATGAACGGTGTTCACGAAGACA	630
152	ThrAspProLeuLysTyrThrLysMetLysGluArgLeuValGlyValSerGluGluThr	171
631	ACAACAGGTGTCACACCGCTCTACAGCTCGAAGAGGAGGCAAACTCTCTTCCACGCC	690
172	ThrThrGlyValLysArgLeuTyrGlnMetGlnAlaAsnGlyThrLeuLeuPheProAla	191
691	ATCAAGCTCAACGCGCTGTACAAAGTCCAAAGTTCGATACATCATCCGGCTCCGCCAC	750
192	IleAsnValAspSerValThrLysSerLysPheAspAsnLeuTyrGlyCysArgHis	211
751	TCCTCTTACGATGGTATCAACCGTGTCTCCGATGTCATGATCGCGCGCAAGACACTCTC	810
212	SerLeuProAspGlyLeuMetArgAlaThrAspValMetIleAlaGlyLysValAlaLeu	231
811	GTATCGGTTACGGCATGTCGCAAGGCTCGCTCAATCCCTCGTGGCCAAAGCCGT	870
232	ValAlaGlyTyrGlyAspValGlyLysGlyCysAlaAlaLeuLysGlnAlaGlyAla	251
871	CGCGTTATCATCACAGAACTCGACCAACTCGCTCGCTCCAGGCTGCCATGGAAGGCTAC	930
252	ArgValIleValThrGluIleAspProIleCysAlaLeuGlnAlaThrMetGluGlyLeu	271
931	CAGTCGCGCGCATCGAAGGAAGTGTCAAGATGTCGATATCTTCGTACATGCACAGA	990
272	GlnValLeuThrLeuGluAspValValSerAspValAspIlePheValThrThrGly	291
991	AACTGCGATATCATCTCTGTTGATCATGCGCCAGATGAAGGATAAGGCTATTGTCGT	1050
292	AsnLysAspIleIleMetValAspHisMetArgLysMetLysAsnAlaIleValCys	311
1051	AACATCGGCACTTCGATPACGAATAATGATACAGATGGCCCTCATGAATACCCAGCATC	1110
312	AsnIleGlyHisPheAspAsnGluIleAspMetLeuGlyLeuGluThrPheProGlyVal	331
1111	AAGCATATCCCAATCAACGCCAGATACGATGTGGGAATCCACAGATGGCCACGCT---	1167
332	LysArgIleThrIleLysProGlnThrAspArgTfpValPheProAspThrAsnSerGly	351
1168	ATCTCTCTCTTCGTGAGGCGCGCTTCTTAACCTTGGCTCGCTACAGGTCAACCATCT	1227
352	IleIleValLeuAlaGluGlyArgLeuMetAsnLeuLeuGlyCysAlaThrGlyHisProSer	371
1228	TTGTTATGTCAATGTCAATTCACAACACGACACTCGCTCGACTCGACCTTACGAAAG	1287
372	PheValMetSerCysSerPheThrAsnGlnValIleAlaGlnLeuGluLeuTrpLysGlu	391
1288	AGA-----GGAAATCTCGAAGCAGGTTTACACACTCCGAAGACATCTCATCAAGAA	1341
392	LysSerThrGlyLysTyrGluLysLysValTyrValLeuProLysHisLeuAspGluLys	411
1342	GTGCTCGCTCCACTCGGATCTCTCGATGTCCACCTTACAAAGCTTACACAGAACGAG	1401
412	ValAlaAlaLeuHisLeuGlyLysLeuGluValAlaArgLeuThrLysLeuSerLysAspGln	431



DE Adenosylhomocysteinase.  
GN AHY.  
OS uncultured; proteobacterium.  
OC Bacteria; Proteobacteria; environmental samples.  
OX NCBI\_TaxID=153809;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21822632; PubMed=11832943;  
RA Beja O., Suzuki M.T., Heidelberg J.F., Nelson W.C., Preston C.M.,  
RA Hamada T., Eisen J.A., Fraser C.M., DeLong E.F.;  
RT "Unsuspected diversity among marine aerobic anoxygenic phototrophs.";  
RL Nature 415:630-633(2002).  
DR EMBL; AE008921; AA048714.1; -.  
DR InterPro; IPR00043; Ado\_hcyase.  
DR InterPro; IPR000205; NAD binding.  
DR Pfam; PF00670; AdoHcyase; 1.  
DR TIGRPFAMs; TIGR00936; ahcy; 1.  
DR PROSITE; P800739; ADHCHYASE 2; 1.  
SQ SEQUENCE 463 AA; 50797 MW; 47A99D102B92D222 CRC64;  
  
Alignment Scores:  
Pred. No.: 6,15e-92 Length: 463  
Score: 1355.50 Matches: 275  
Percent Similarity: 70.56% Conservative: 63  
Best Local Similarity: 57.41% Mismatches: 118  
Query Match: 49.53% Indels: 23  
DB: 2 Gaps: 5  
  
US-09-759-990A-1 (1-1461) x Q8XYV9 (1-463)  
QY 34 GAGTACAGAAATTCGGCAGTACCACTCCATGTTCTCGCGGTAAAGAACTTACCTTGCT 93  
Db 4 AspyrTlleValLysAspIleSerLeuAlaAlaPheGlyArgLysGluLeuAlaAla 23  
QY 94 GAGAGGAATCCAGAGTCTTATGTTCTGTCGAGCGTTATTCGCTTCTAAGCAATG 153  
Db 24 GluThrGluMetProGlyLeuMetAlaLeuArgLysGluTyrGlyAlaAlaLysProLeu 43  
QY 154 AAGGTGTCAGAACTCTGTTCTCCATGATGATGATGATGATGATGATGATGATGATGAT 213  
Db 44 LysGlyAlaArgIleValGlySerLeuHisMetThrIleGlnThrAlaValLeuIleGlu 63  
QY 214 ACATCCAGATCTTGGTGTGATGTCAGATGGGCTTCCTGCAACATCTTCTACACAA 273  
Db 64 ThrLeuValGluLeuGlyAlaAspValArgTrpAlaSerCysAsnIlePheSerThrGln 83  
QY 274 GATACAGCGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 333  
Db 84 AspHis\*\*\*Ala\*\*\*AlaIleAlaAlaGly-----Gly 94  
QY 334 ATCCAGATCTTCGCTGAGGGGCAACACTCCAGAACTCTCCAGAACTCTCCAGAACTCT 393  
Db 95 ThrProValPheAlaIleLysGlyGlnSerLeuGlnHisTrpAspTyrLeuAspArg 114  
QY 394 GCTCTCAGATGCGCAGATGCTCAAGGCCCAACAGCAGGTTCGATGATGCTGATGCT 453  
Db 115 SerPheMetPhePro-----GluGlyAlaAsnMetIleLeuAspGlyGlyAspAla 132  
QY 454 ACATCTCTCACTCCAGAGGCTCGAATTCGAACTCGAACTCGAACTCGAACTCGAACT 501  
Db 133 ThrLeuTyrValLeuLeuGlyAlaArgValGluGlyGluThrAspLeuIleAlaVal 152  
QY 502 CCAGAGCCAAAGAGTGTCAACACTCGAATACCGCTGCTTCTGCTACCTCAAGCAG 561  
Db 153 ProLysSerGluGluGlu-----AlaValPheAlaGlnIleLysLys 167  
QY 562 GTCTTCAACAGACAGAACCTCGGCACAGTTCGTCGCGCATGACCGTGTTC 621  
Db 168 ArgValAlaGluThrProGlyTrpPheThrLysThrArgAlaAlaIleGlnGlyValSer 187  
QY 622 GAGAGACACACAGAGGTGCTCCAGCGCTCTACAGCTCGAGAGGAGGCGCAACTCTC 681  
Db 188 GluGluThrThrThrGlyValHisArgLeuTyrGluLeuValLysGlnGlyGlnLeuPro 207

QY 682 TTCCAGCCATCAACGTCAACAGCGCTGTGTACAAAGTCAAAGTTCGATAACATCTACGC 741  
Db 208 PheProAlaIleAsnValAsnAspSerValThrLysSerLysPheAspAsnLysTyrGly 227  
QY 742 TGGCGCCATCTCCCTTATCGATGGTATCAACCGTGTCTCCGATGTGTCATGATCGCGCGCAG 801  
Db 228 CysIysGluSerLeuValAspGlyIleArgArgAlaThrAspThrMetMetAlaGlyLys 247  
QY 802 ACAGCTCTGCTCATGGTTACGGCGATGTCGGCAAGGCTCGCTCAATCCCTCCCTGGC 861  
Db 248 ValAlaValIleValGlyTyrGlyAspValGlyLysGlySerAlaAlaSerLeuArgGly 267  
QY 862 CAAGCGCTCGCTTATCATCACAGAACTCGACCCCAATCTCGCTCTCCAGGTGCGCATG 921  
Db 268 AlaGlyAlaArgValLysValThrGluValAspProIleCysAlaLeuGlnAlaAlaMet 287  
QY 922 GAAGCTACCGAGTCCGCGCATCGAGGAAGTCGCAAGGATGTCGATATCTCGTTACA 981  
Db 288 AspGlyPheGluValValLeuLeuGluAspValValSerSerAlaAspIlePheIleThr 307  
QY 982 TSCACAGGAACTCGCATATCATCTCTGTTGACATGATGCCGCCAGATGAAGGATAGGCT 1041  
Db 308 ThrThrGlyAsnLysAspValIleArgIleGluHisMetArgAspMetLysAspMetAla 327  
QY 1042 ATTGTCGTGATCATCGGCCACTTCGATACGAAATTTGATACAGATGGCCTCATGAATAC 1101  
Db 328 IleValGlyAsnIleGlyHisPheAspAsnGluIleGlnValAlaSerLeuLysAsnHis 347  
QY 1102 CCAGGCATCAAGCACATCCCAATCAAGCCAGATGACATGCGGAATTCGCCAGATGGC 1161  
Db 348 -----LysTrpThrAsnIleLysGluGlnValAspMetIleGluMetProAsnGly 364  
QY 1162 CACGCTATCTCTCTGTCGAGGGCGCTTCTTAACCTTGGCTGGCTGCTACAGTCTAC 1221  
Db 365 AsnArgIleIleLeuLeuSerGlyArgLeuLeuAsnLeuGlyAsnAlaThrGlyHis 384  
QY 1222 CCATCTTCTGTTATGTCATGTCATTCACAAACACAGACATCGCTCAGCTCGACCTCTAC 1281  
Db 385 ProSerPheValMetSerAlaSerPheThrAsnGlnValLeuAlaGlnIleGluLeuTrp 404  
QY 1282 GAAAGAGAGGAAATCTCGAGAAGAGGTTTACACTTCCGAAGCATCTCGATGAAGAA 1341  
Db 405 ThrLysGlyAspAspTyrSerAsnGluValTyrIleLeuProLysHisLeuAspGluLys 424  
QY 1342 GTCCCTCGCTCCACTCGATCTCTCGATGTCACCTTACAACTTACAACTTACAAAGAGCAG 1401  
Db 425 ValAlaArgLeuHisLeuAspArgIleGlyValLysLeuSerLysLeuAsnAspGluGln 444  
QY 1402 GCTGACTACATCAACGTTCCAGTTGAGGTCCTTACAGTCTGATGCTTACCGTTAT 1458  
Db 445 AlaAlaTyrIleGlyValSerSerGluGlyProPheLysProGluHisTyrArgTyr 463  
  
RESULT 14  
Q8XYV1 PRELIMINARY; PRT; 436 AA.  
AC Q8XYV1;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE S-adenosylhomocysteine hydrolase.  
GN LB106.  
OS Leptospira interrogans.  
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
OX NCBI\_TaxID=173;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Seroovar lai;  
RA Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE011599; AA051665.1; -.  
KW Hydrolase; Complete proteome.  
SQ SEQUENCE 436 AA; 48233 MW; 1E73AE26EBDFAC1 CRC64;

Alignment Scores:				
Pred. No.:	1,19e-91	Length:	436	
Score:	1351.50	Matches:	272	
Best Local Similarity:	69.62%	Conservative:	58	
Query Match:	49.38%	Mismatches:	95	
DB:	16	Indels:	5	
US-09-759-990A-1 (1-1461) x Q8EXV1 (1-436)				
QY	37	TACAGATTGCGGACATCAACCTCCATGTTCTCGGCGGTGAAGAACTTACCTTCGTCGAG	96	
DB	12	TyrlsValLysAspLeuSerGlnAlaGluTrpGlyArgGlnGluIleLeuAlaGlu	31	
QY	97	AAGCAATCCAGAGTCTTATGTTCTTCGTGAGCGTATTCCGCTTCTAAGCCATTGAAG	156	
DB	32	LysGluMetProGlyLeuMetAlaLeuArgGlnGluTrpLysGlyLysLysProLeuAla	51	
QY	157	GGTGTCAAGATCTGTGTTCCCTCCATGACAGTCCAGACAGCGTCTCATGACAGACA	216	
DB	52	GlyAlaArgIleAlaGlySerLeuHisMetThrIleGlnThrAlaValLeuIleGluThr	71	
QY	217	CTCACAGCTTGTGTGTGATGTCAGATGGGCTTCCTGCAACATCTTCTTACACAGAT	276	
DB	72	LeuThrGluLeuGlyAlaGluValArgTrpSerCysAsnIlePheSerThrGlnAsp	91	
QY	277	ACAGCGCTGTGCTGTATGTTGTGCGCCCAACAGSCACACAGAGAAGCCAGCGGTATC	336	
DB	92	HisAlaAlaAlaIleAla-----LysAlaGlyIle	102	
QY	337	CGAGTCTTCGCTGAGGCGGAAACACTCCCGAATCTGGAGAACACATACCGCGCT	396	
DB	103	ProValPheAlaTrpLysGlyGluThrGluGluGluTrpTrpCysIleGluGlnThr	122	
QY	397	CTCACATGCCAGATGTTCAAGGCCCAACAGCAGGTGTGCGATGCTGGTGTATCTACA	456	
DB	123	IlePhePheGlyAsp--LysGlyProAsnMetIleLeuAspAspGlyAspLeuThr	141	
QY	457	CTCCTCATCTCAGGCGTTCGAATTCGAAACAGCGGTGTGTTCTCAGAGCCACAGAA	516	
DB	142	AlaTyrlleHisGlu-----	146	
QY	517	GCTGACACCTCGAATACCGTGTCTTGTCTGATCAACAGCAGGTCTTCAACAGAC	576	
DB	147	-----LysTyrlProLysLeuLeuSerGluIleArg-----	156	
QY	577	AAGAACCTGCGCACACAGTGTGCGCGCATGAACCGTGTTCGGAAGACACACAA	636	
DB	157	-----GlyIleSerGluGluThrThrThr	164	
QY	637	GGTGTCCACCGCTCTACAGCTCGAAGAGGAGCAACTCTCTTCCACGACCATCAAC	696	
DB	165	GlyValLysSerLeuTyrlsLeuLeuLysLysGlyGluLeuLysValProIlePheAsn	184	
QY	697	GTCACAGCGTGTTCAGAACTCAAGTTCGATACACTACGCGTCCGCCCATCCCTT	756	
DB	185	ValAsnAspSerValThrLysSerLysPheAspAsnLeuTyrlGlyCysArgGluSerLeu	204	
QY	757	ATCGATGATCAACCGTGTCTCGATGTCATGTCGGCGGACAGACACTCTCTGTCATG	816	
DB	205	AlaAspGlyIleLysArgAlaThrAspValMetLeuAlaGlyLysValAlaLeuValCys	224	
QY	817	GGTGTAGCGGATGTCGAGGCGTGGCTCAATCCCTCCGTGCGCAAGCGCTCGCGTT	876	
DB	225	GlyPheGlyAspValGlyLysGlySerAlaAlaSerLeuArgAsnPheGlyAlaArgVal	244	
QY	877	ATCATCACAGACTCGACCCCAATCTGCTTCCAGCGTCCATGAGAGGCTACAGATC	936	
DB	245	IleValThrGluIleAspProIleCysAlaLeuGlnAlaSerMetGluGlyTyrlGlnVal	264	
QY	937	CGCGCATCGAGAACTGTCAGGATGTCGATATCTTCTGATCATGTCAGAGAACTGC	996	

Db	265	LeuArgValGluAspIleIleGluGlnValAspIleValThrAlaThrGlyAsnAsp	284	
QY	997	GATATCATCTCTGTGACATGATGCGCCACAGGATAGGCTATTGTCGGTACATC	1056	
Db	285	AspIleIleThrLeuGluHisMetLysAlaMetLysAspGlyAlaIleLeuCysAsnIle	304	
QY	1057	GGCCACTTCGATACCAAAATGATACAGATGCGCTCATCAAAATACCCAGGCATCAAGCAC	1116	
Db	305	GlyHisPheAspThrGluIleGlnMetSerArgLeuAsnGlnLysGlyValThrLys	324	
QY	1117	ATCCCATCAAGCCAGATAACGACATGTGGAAATCCAGATGCGCCACGCTATCTCCTT	1176	
Db	325	LysGluIleLysProGlnValAspLysTyThrPheProAspGlyLysSerIleIleVal	344	
QY	1177	CTTCTCAGGCGCCCTCTTAACTCTGGCTGCGCTACAGGTCCACCATCTTTCTGTTATG	1236	
Db	345	LeuAlaGluGlyArgLeuValAsnLeuGlyCysAlaThrGlyHisProSerPheValMet	364	
QY	1237	TCAATGTCAATCAAAACACAGACACTCGCTCAGCTCGACCTCTACGAAAGAGAGAAAT	1296	
Db	365	SerCysSerPheThrAsnGlnValLeuAlaGlnIleGluLeuTyrlAsnAsnLys-----	382	
QY	1297	CTCGAGAAGAGTTTACACACTTCCGAGCATCTCGATGAAGAGTCTGCTCCCTCCAC	1356	
Db	383	TyrlGluLeuGlyValTyThrLeuProLysHisLeuAspGlyLysValAlaAlaLeuHis	402	
QY	1357	CTCGATCTCTCGATGCCACTTCAAACTTACACAGAGCAGGCTGACTACATCAAC	1416	
Db	403	LeuGluGlnLeuGlyValArgLeuThrLysLeuAsnGlnLysGlnAlaAspTyrlLeuGly	422	
QY	1417	GTTCCAGTTCAGGCTCTTCAAGTCTGATGCTTACCGTTAT	1458	
Db	423	ValProIleAsnGlyProPheLysProAspHisTyrlArgTyrl	436	
RESULT 15				
ID	Q8FXZ7	PRELIMINARY;	PRT;	466 AA.
AC	Q8FXZ7			
DT	01-WAR-2003	(TrEMBLrel. 23, Created)		
DT	01-WAR-2003	(TrEMBLrel. 23, Last sequence update)		
DE	01-WAR-2003	(TrEMBLrel. 23, Last annotation update)		
DE	Adenosylhomocysteinase.			
GN	AHCY OR BR2097.			
OS	Brucella suis.			
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;			
OC	Brucellaceae; Brucella.			
OX	NCBI_TaxID=29461;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=1330 / Biovar 1;			
RC	MEDLINE=2224774; PubMed=12271122;			
RA	Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,			
RA	Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,			
RA	Laughtery S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,			
RA	Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,			
RA	Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,			
RA	Hoover D.L., Lindler L.E., Hailing S.M., Boyle S.M., Fraser C.M.;			
RT	"The Brucella suis genome reveals fundamental similarities between			
RT	animal and plant pathogens and symbionts.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153 (2002).			
DR	EMBL; AE014496; AAN30987.1; -.			
DR	TIGR; BR2097; -.			
KW	Complete proteome.			
SQ	SEQUENCE 466 AA; 50791 MW; 6AB3B4B1C2F0BE41 CRC64;			
Alignment Scores:				
Pred. No.:	3.12e-91	Length:	466	
Score:	1346.00	Matches:	272	
Percent Similarity:	70.74%	Conservative:	64	
Best Local Similarity:	57.26%	Mismatches:	125	
Query Match:	49.18%	Indels:	14	
DB:	16	Gaps:	4	

US-09-759-990A-1 (1-1461) x Q8FXZ7 (1-466)

```
QY 34 GAGTACAAATTCGCGACATCAACTCCATGTTCTCGCGGTAAAGAACTTACCCCTTGCT 93
Db 6 AspPheValValLysAspIleSerLeuAlaAspTrpGlyArgLysGluLeuAla 25
QY 94 GAGAAGAAATTCGCGATCTTATGTTCTTCGTGAGCGTTATTCCGCTTCTAAAGCCATTG 153
Db 26 GluThrGluMetProGlyLeuMetAlaAlaArgGluLeuPheGlyLysSerGlnProLeu 45
QY 154 AAGGTGTCAAGATCTGTTCCCTCCACATGACATCCAGACAGCGTCTCATCGAG 213
Db 46 LysGlyAlaArgIleSerGlySerLeuHisMetThrIleGlnThrAlaValLeuIleGlu 65
QY 214 ACATCCACAGCTCTTGCTGATGTCAGATGGGCTTCGCAACATCTTCTTACACAA 273
Db 66 ThrLeuLysValLeuGlyAlaGluValArgTrpAlaSerCysAsnIlePheSerThrGln 85
QY 274 GATACAGCGCTGCTGCTATCGTTGCTGGGCCCAACAGACACACCCAGAGAGCCGCGGT 333
Db 86 AspHisAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 97
QY 334 ATCCCAAGCTTCGCTCGAAGGCGGAACACTCCACAGATCTCCGAGAGACACATACCGC 393
Db 98 ---ProValPheAlaValLysGlyGluThrLeuGluGluIleTrpThrTyrThrAspGln 116
QY 394 GCTCTCACATGCGCAGATGTCAGGCCCCACAGCAGGTTGTCGATGATGGTGTGATGCT 453
Db 117 IlePheGlnTrpProAspGlyGluProSerAsnMetIleLeuAspAspGlyGlyAspAla 136
QY 454 ACATCTCTCATCTCCAGGCTTCGAAATTCGAAGACACCGGCTGCTCCAGAGCCCAACA 513
Db 137 ThrMetTyrIleLeuIleGlyAlaArgAlaGluAlaGluAlaGluAspValLeuSerAsnPro 156
QY 514 GAAGTCGAACCTCGAATACCGCTGCTTCTGCTACTCAAGCAGGCTTCTCAACCAA 573
Db 157 GlnSerGlu-----GluGluGluValLeuPheAlaGlnIleLysLysArgMetAlaAla 174
QY 574 GACAGAACCTACGCGACAGATGCTGCGCGCATGACGGTGTTCGAGAGACACAA 633
Db 175 ThrProGlyPhePheThrLysGlnArgAlaAlaIleLysGlyValThrGluGluThr 194
QY 634 ACAGGTGTCACCGCTCTACAGCTCGAGAGAGAGGCAAACTCCTCTTCCAGCCATC 693
Db 195 ThrGlyValAsnArgLeuTyrGlnLeuGlnLysLysGlyLeuLeuProPheProAlaIle 214
QY 694 AACGTCAACGCGCTGTTACAAAGTCCAAAGTTCGATACATCTACGCTGCGCGCACTCC 753
Db 215 AsnValAsnAspSerValThrLysSerLysPheAspAsnLysTyrGlyCysLysGluSer 234
QY 754 CTTATCGATGTTATCAACCGCTCTCCGATGTCATGATCGCGCGCAAGACAGCTCTGTC 813
Db 235 LeuValAspGlyIleArgArgGlyThrAspValMetMetAlaGlyLysValAlaValVal 254
QY 814 ATGGGTACGGCGATGTCGCAAGGCTCGCTCAATCCCTCCGTCGCGCAAGGCGCTGCG 873
Db 255 CysGlyTyrGlyAspValGlyLysGlySerAlaGlnSerLeuAlaGlyAlaGlyAlaArg 274
QY 874 GTTATCATCAGAACTCGACCAATCTCGCTCTCCAGGCTGCGCATGGAAGGTACCAG 933
Db 275 ValLysValThrGluValAspProIleCysAlaLeuGlnAlaAlaMetAspGlyPheGlu 294
QY 934 GTCGCGCATCGAGAAAGTCGTCAGGATGTCGATATCTTCGTTTACATGCACAGGAAAC 993
Db 295 ValValThrLeuAspAspAlaAlaSerThrAlaAspIleValValThrThrGlyAsn 314
QY 994 TCGCATATCATCTCTGTGACATGATGCGCCCGATGAGAGGATAGGCTATTCTCGTTAC 1053
Db 315 LysAspValIleThrIleAspHisMetArgLysMetLysAspMetCysIleValGlyAsn 334
QY 1054 ATCGGCCCTCTCGATAACGAAATGATACAGATGGCTCATGAAATACCCAGCATCAAG 1113
Db 335 IleGlyHisPheAsnGluIleGlnValAlaAlaLeuArg-----AsnLeuLys 351
```

```
QY 1114 CACATCCCAATCAAGCCAGAAATACGATGTCGGAAATTCACGATGGCCACGCTATCCTC 1173
Db 352 TrpThrAsnValLysProGlnValAspLeuIleGluPheProAspGlyLysArgLeuIle 371
QY 1174 CTTCTTGCTGAGGCGCGCTTCTTAACCTTGGCTCGCTACAGGTACCCATCTTTCGTT 1233
Db 372 LeuLeuSerGluGlyArgLeuLeuAsnLeuGlyAsnAlaThrGlyHisProSerPheVal 391
QY 1234 ATGTCAATGTCATTCCAAACCCAGACACTCGCTCAGCTCGACCTCTACGAAAGAGAGGA 1293
Db 392 MetSerAlaSerPheThrAsnGlnValLeuGlyGlnIleGluLeuPheThrArgThrAsp 411
QY 1294 AATCTCGAAGAAAGAGTTTACACACTTCCGAAAGCATCTCGATGAAGAGTCGCTCGCTC 1353
Db 412 AlaTyrLysAsnGluValTyrValLeuProLysHisLeuAspGluLysValAlaArgLeu 431
QY 1354 CACCTCGGATCTCTCGATGTCACCTTACAAAGCTTACACAGAGCAGGCTGACTACATC 1413
Db 432 HisLeuAspLysLeuGlyAlaLysLeuThrValLeuSerGluGlnAlaAlaIleTyrIle 451
QY 1414 AACGTTCCAGTTGAGGGTCCCTTACAGTCTGATGCTTACCGTTAT 1458
Db 452 GlyValThrProGlnGlyProPheLysSerGluHisTyrArgTyr 466
```

Search completed: October 15, 2003, 14:53:27  
Job time : 183 secs